Working Paper Series

The Complexity of CEO Compensation

WP 14-16

Arantxa Jarque
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

The Complexity of CEO Compensation*

Arantxa Jarque[†]
Richmond Fed

September 3, 2014
Working Paper No. 14-16

Abstract

I study firm characteristics that justify the use of options or refresher grants in the optimal compensation packages for CEOs in the presence of moral hazard. I model explicitly the determination of stock prices as a function of the output realizations of the firm: Symmetric learning by all parties about the exogenous quality of the firm makes stock prices sensitive to output observations. Compensation packages are designed to transform this sensitivity of prices–to–output into the sensitivity of consumption–to–output that is dictated by the optimal contract. Heterogeneity in the structure of firm uncertainty implies that some firms are able to implement the optimal contract with very simple schemes that do not include options, refresher grants, or perks, while others need to use these more complex and potentially less transparent instruments.

Journal of Economic Literature Classification Numbers: D80, D82, D86, G30.

Key Words: mechanism design; moral hazard; CEO compensation; stock options; repricing; refresher grants; perks; learning

1 Introduction

It is widely accepted that in order to solve the agency problem between a firm's CEO and its owners the compensation of the executive must be tied to the results of the firm. Less well understood, however, is how the efficient provision of incentives must be implemented in practice. The recent financial crisis has revived a longstanding debate about compensation practices for CEOs of large, publicly traded firms. Although the interest in explaining the level of compensation is still present,[1]

*I would like to thank Andy Atkeson, Marco Celentani, Hector Chade, Huberto Ennis, Willie Fuchs, Juan Carlos Hatchondo, Ángel Hernando, Hugo Hopenhayn, Boyan Jovanovic, Nobu Kiyotaki, Ned Prescott, and audiences at the University of Iowa, the Feedback group at Carlos III, Yonsei University, the Applied System meetings and the Midwest Macro Meetings in Spring 2014, and the SED meetings in Toronto for valuable comments. I also thank Robert Sharp for excellent research assistance.

[†]Correspondence: Arantxa.Jarque@rich.frb.org. The views expressed in this paper are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Richmond or those of the Federal Reserve System.

[1]Academics have recently proposed explanations for the increase of the level of pay in the past decades, mainly based on assortative matching combined with a sharp increase in the size of firms during this period (see Gabaix and Landier (2008) and references therein).

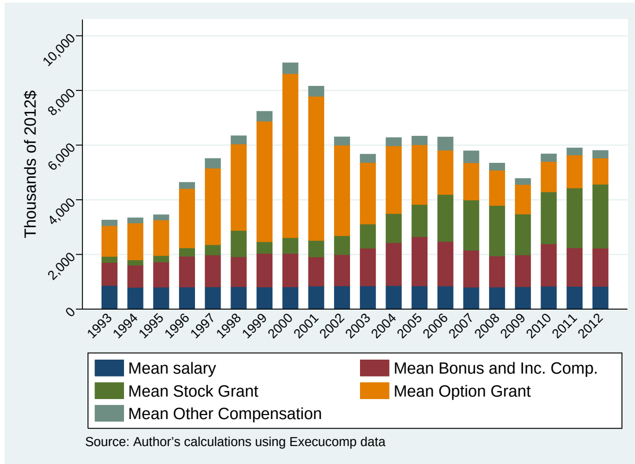

Figure 1: Relative importance of the different components of CEO compensation packages over time. Column height represents average compensation. Author's calculations using Execucomp data (CEOs of the largest 1,500 firms that are listed in the S&P index). CEOs who owned more than 3% of the total stock of the firm at any point in the sample were considered "owners," i.e., not subject to a moral hazard problem, and they were dropped.

there has been a shift toward understanding the *form* of compensation packages: Is the use of certain instruments, like stock options, golden shakehands and parachutes, or perks, a sign of captured compensation boards and misaligned incentives?[2] In a similar spirit, concerns that certain pay practices, like the use of options, may induce excessive risk–taking in financial firms have prompted regulatory agencies to increase their involvement in overseeing pay practices in the banking sector.[3] Given the state of the debate and the increased desire for intervention, it is critical to enhance our understanding of the form of compensation packages that *are* consistent with a correct alignment of the interests of shareholders and the CEO. This is the objective of this paper.

Contract theory informs us about how to implement incentives optimally (Holmstrom, 1979; Grossman and Hart, 1983; Wang, 1997). But such characterizations are mainly given in a context that allows the payoff to the agent to depend on signals of performance (e.g., accounting measures, stock prices) in a very general way (i.e., "unrestricted" transfers). In this paper, I take a different approach: I look at the problem of implementing the optimal contract with a rich set of real–life instruments, and I study what are the necessary instruments to implement it exactly.

This approach differentiates my paper from important recent contributions in the literature that studies the implementation of incentives with real–life instruments. Available data on the

[2] See Bebchuk, Cohen, and Spamann (2010) and references therein for an elaboration of this argument. Fahlenbrach and Stulz (2009), however, find no support for it in their data.

[3] See, for example, the 6/10/2009 statement by then Treasury Secretary Tim Geithner on executive compensation in which he provided broad–based principles as a first step in "the process of bringing compensation practices more tightly in line with the interests of shareholders and reinforcing the stability of firms and the financial system." The statement is available at http://www.treasury.gov/press-center/press-releases/Pages/tg163.aspx.

compensation of CEOs of large, publicly traded firms (Execucomp) shows that a fairly limited set of compensation instruments, like bonus programs and stock and option grants, make up most of CEO compensation (see Figure 1).[4] Perhaps because of this evidence, a common practice in the literature that studies the properties of real life compensation instruments has been to exogenously restrict the class of instruments that are available to the firm and derive the optimal scheme within that class. See, for example, Clementi, Cooley, and Wang (2006), Kadan and Swinkles (2007), or Edmans and Liu (2010). The restriction in these studies on the number and generality of instruments is necessary for practical purposes, since the complexity of the optimization problem increases very fast with the number of instruments allowed. Although this approach yields interesting insights on the way particular instruments may work, it poses a potentially important shortcoming: The results rely on restricting the firm to use an inefficient compensation scheme.[5] This is not the case in my setup, since I allow for a rich enough set of instruments so that the transfers given to the CEO with the compensation package correspond to those in the optimal contract derived assuming an unrestricted set of instruments. The issue that I want to study is not whether, for an exogenously given set of compensation instruments, the optimal contract is feasible, or how closely it can be approximated, but rather what are the necessary instruments to implement it exactly.

My model shows how a firm's unobserved characteristics (related to the uncertainty about the moral hazard problem it faces) may affect the "complexity" of its compensation package. For example, I show that, for some firms, the use of options will be necessary to implement the optimal contract, while it will be optional for others. This is consistent with existing empirical evidence of cross–sectional variation in compensation practices. Figure 2 presents publicly available data about the complexity of compensation packages of CEOs of large, publicly traded firms (Execucomp).[6] It plots the evolution over time of the percentage of firms that include in the compensation package of their CEO in a given year the following groups of instruments: both stock and options grants ($I = B$), only stock ($I = S$), only options ($I = O$), or neither stock nor options ($I = N$).[7] Two facts stand out. First, in any given year there is heterogeneity in the instruments used by firms. Second, the introduction of mandatory expensing of at–the–money options around 2006 that arguably made options less attractive for accounting purposes seems to have pushed some firms to stop using them, but after the change there is still an important fraction of firms using options. This provides further evidence for heterogeneity in the cross–section in the value of including options in compensation schemes.[8] My model provides a justification for (unexplained) heterogeneity in the composition of

[4]A reason frequently cited for the proliferation of option grants is the tax advantages of compensation contingent on performance over base salaries. The Omnibus Budget Reconciliation Act (OBRA) resolution 162(m) of 1992 imposed a $1 million cap on the amount of non–performance–based compensation of the top executives of the firm that qualifies for a tax deduction. Certain restricted stock and option plans are considered performance–based pay.

[5]Two important exceptions, which I discuss later in this introduction, are Aseff and Santos (2005) and Edmans et al. (2012).

[6]See also Kole (1997), Kadan and Swinkles (2007), and references therein for previous studies exploring empirically the relationship between pay features (like the choice of stock versus options or the importance of incentive pay) with the firm's observable characteristics (such as size, industry, age, or default risk).

[7]These are all firms for which the CEO owns less than 3% of the stock and which have a positive number (regardless of its magnitude) in either stock or option grants in the Execucomp entry for the compensation of their CEO in a given year.

[8]Moreover, the new accounting standards for expensing options that were only mandatory in 2006 were voluntarily

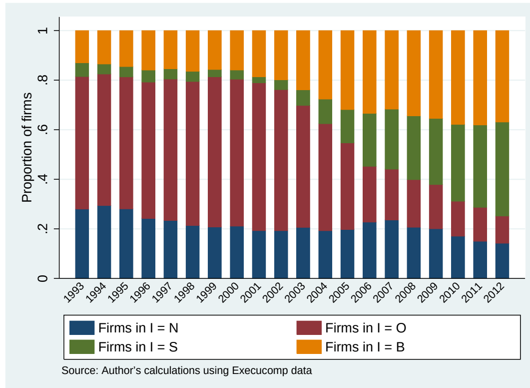

Figure 2: Percentage of firms that include in the compensation package of their CEO in a given year the following groups of instruments: both stock and options grants ($I = B$), only stock ($I = S$), only options ($I = O$), or neither stock nor options ($I = N$) .

compensation packages.[9]

I model the moral hazard problem between the owners of the firm and the CEO as a principal–agent problem. I propose a simple two–period framework in which a risk–averse CEO is asked to exert an unobservable and costly effort in the first period only. The risk–neutral owners of the firm coordinate to act as the unique principal who designs the compensation package of the CEO. I assume commitment to the two–period contract for both the CEO and the firm owners, and I abstract from firing or quitting decisions. The only effort of the CEO determines the distribution of the firm's output in both periods. I interpret the first period as an interim stage at which information about performance is revealed (the company announces its earnings in the middle of the fiscal year). New grants may be awarded to the CEO at this interim stage (refresher grants), but no consumption takes place then; the CEO receives and consumes his compensation in the second period only, after two output realizations have been observed.

With this timing I try to capture the fact that relevant information about performance is typically revealed between the time at which compensation packages are set and the time when the CEO collects his payments. Hence, the expected payouts from bonus plans and stock or option grants may change with interim earnings announcements. In the simplified version of my framework without learning, however, when the firm is valued according to the expected stream of future output, stock prices do not change contingent on the value of earnings announced in the interim stage. This is because, in equilibrium, the recommended level of effort is chosen, and that alone determines the distribution of output. Hence, the expectations about output in the second period are independent of the first period realization. In reality, however, stock prices typically change over the course of a fiscal year with the earnings announcements of the firm. To capture

adhered to by some firms starting in 2002 (Carter, Lynch, and Tuna, 2007).

[9] Jarque and Gaines (2013), for example, find that much of this cross–sectional heterogeneity remains unexplained after controlling for firm characteristics such as size and the level of compensation, as well as for the 2006 changes in the mandatory expensing of options granted at the money.

this feature, I augment the model by introducing an exogenous source of uncertainty: a stochastic state that affects the effectiveness of the effort of the CEO.[10] This can be interpreted as the quality of the match between the CEO and the firm, or as idiosyncratic market conditions. I assume that both the CEO and the owners, as well as buyers of the stock, have a prior about this state, which they update through Bayes' rule when they observe a new realization of output. This generates contingent stock prices.

In my full model with learning, then, the distribution of stock prices contingent on effort is endogenously generated through the learning process, rather than assumed as a primitive. Importantly, this implies that consumption in the optimal contract need not be monotonic in output, even though effort first–order stochastically increases output in each of the states. This is a key difference with most of the literature, and in particular with Aseff and Santos (2005), who characterize the optimal contract in a standard static moral hazard problem (Grossman and Hart, 1986) under the assumption that effort affects directly the distribution over stock prices. When they calculate numerically the cost of limiting the compensation scheme to include only a wage and an option grant, their calibrated model using CEO compensation data implies small costs for the approximation. However, their assumptions imply contracts that are always monotonic in the stock price, and hence their result is not directly applicable to my setup.

When studying the implementation of the optimal contract with real–life compensation instruments, the stylized compensation package that I consider tries to include all the standard instruments that we observe in real life (a salary, a bonus program, stock grants, and option grants), as well as less standard but often observed types of compensation such as signing bonuses, severance payments, perks, discounted stock purchases, or below–market–rate loans, which I will refer to generally as "ad–hoc" payments. An important characteristic of the standard instruments is that there is a clear tie between their payoff and the performance of the firm. This may facilitate both the communication of stockholders' objectives to CEOs, and the transparency of compensation practices to potential outside investors and regulators. The defining characteristic of ad–hoc payments, instead, will be that, even though they are understood and anticipated by the CEO, they cannot be calculated by an outside observer as a function of annual accounting measures or the stock price realization. That is, even though these transfers may follow from a (non–formulaic, potentially subjective) evaluation of the performance of the CEO using accounting and stock price realizations, to an outside observer these remain arbitrary transfers of money.[11]

[10] This is a simple way to make prices respond to output realizations. One could generalize the framework to allow also for output persistence; even though this (or any other alternative stock pricing rule) would likely affect the particular characterizations in the core of the paper (and the tractability), the main idea that compensation packages need to translate the sensitivity of stock–to–output into optimal consumption–to–output is robust to such alternative specifications.

[11] Identifying ad–hoc payments in the Execucomp data is not easy, since some of the "standard" compensation instruments (like stock and option grants) that we observe in compensation plans may also be granted in an "ad–hoc" manner, and this difference is not apparent in Execucomp. Gillan, Hartzell, and Parrino (2009) presents han–collected evidence that slightly less than half of the firms in the S&P 500 in the year 2000 had explicit Employment Agreements describing compensation agreements. For those that had them, the agreement covered on average three years, and it specified mostly the initial salary, potential reasons for dismissal, and the compensation that would folow in such events, as well as in the event of a change in corporate control. About half of the contracts specified a bonus target,

To present my conclusions about the form of real–life compensation packages, I define three types of pay schemes, according to the instruments that are included in each. The first distinction is between schemes that include ad–hoc payments and those that do not; I label schemes with ad–hoc payments as "nontransparent," trying to capture with this language the difficulty that I just described for an outsider to understand the link of these payments to performance. In contrast, "transparent" schemes may include only a wage and instruments that depend on output or stock prices. Within the transparent category, I distinguish between "simple" schemes (which may include only stock granted before any realization of output is observed,) or "complex" (which may include also stock options, and stock and option refresher grants issued contingent on the first period realization). Complexity, then, captures not only the number of different instruments but also their sophistication.

My results are the following. First, I show analytically that ad–hoc payments are only necessary for firms whose optimal contract implies consumption that is nonmonotonic in cumulative output. My characterization shows that there is a nontrivial set of firms for which this will be the case. On the other hand, I show analytically that when the optimal contract is monotonic in output, a complex scheme is always sufficient, and it only needs to include a wage, a bonus plan, and, importantly, an option grant that is issued contingent on the interim output realization. Hence, my framework provides a rationale for the seemingly counterintuitive compensation practice of issuing refresher grants. Second, I show numerically that — perhaps surprisingly — simple schemes are sufficient for many types of firms. Finally, I show that there is a nonempty set of firms for which a simple scheme is not sufficient but a complex one is, making refresher stock and option grants instrumental in avoiding ad–hoc compensation.

The result that refresher grants are useful in implementing the optimal contract is due to the fact that the contract may call for different sensitivity of pay–to–output depending on the interim results of the firm, and instruments like bonus plans or plain stock awards cannot always implement this contingent sensitivity. In particular, prices may change very little with output realizations in the second period if the first output was very informative about the exogenous state (i.e., the sensitivity of prices–to–output is very different from the sensitivity of pay–to–output that the optimal contract calls for). To illustrate this, I discuss examples in which awarding new options to a CEO, even after observing bad results, may be part of the implementation of the optimal contract for firms that want to avoid the use of ad–hoc payments. What may superficially look like undoing incentives, or a sign of entrenchment, actually arises as part of the optimal provision of incentives.

The role of refresher grants that I underline in this paper is related to the results in three previous studies. The first one is Kadan and Swinkles (2007). In their setting, compensation instruments are restricted to a wage and an option with an exercise price that is derived optimally. Although their model is static, they model and explicitly study the role of previous outstanding grants, which imply a lower bound on the compensation that the CEO gets for every current price realization. They show that only for financially distressed firms or start–ups (firms with

and details on the retirement plan, and about a third discussed initial restricted stock or option grants, as well as perks (such as car or plane use, or a club membership). For those firms that relied on implicit agreements instead, they show that incentive pay is nontrivial, averaging at 54% of total pay.

substantial nonviability risk) the optimal exercise price of the new (refresher) grants is zero, i.e., stock is a better incentive instrument than options. However, for firms in good financial health, refresher option grants (positive exercise price) help implement a higher sensitivity of payments to performance for the relevant range of market prices (the prices that are likely to be realized going forward given the good financial health of the firm). My results reinforce their conclusions by allowing for a richer set of compensation instruments that does not restrict the form of the contract. A crucial contribution of this paper relative to Kadan and Swinkles (2007) is to have endogenous stock prices, which helps understand an added difficulty in the use of compensation instruments that are contingent on those prices: the possibility of non–monotonicities in the optimal compensation contract.

A second related paper is Acharya, John, and Sundaram (2000). The authors study the practice of repricing options (i.e., changing the exercise price of options previously granted, typically to make options that are currently out–of–the–money be at–the–money). When outstanding options have an exercise price that is viewed as unattainable before the expiration of the grant, repricing is, in practice, equivalent to issuing refresher grants.[12] Their two–period problem reduces formally to a repeated moral hazard with consumption in the final period only. Because they restrict the compensation instruments available to the principal (in particular, refresher grants are not allowed), after some realizations of the interim period output the agent cannot be incentivized to exert (otherwise efficient) effort. Hence, allowing for repricing effectively expands the set of compensation instruments in their setting, which can be useful to provide incentives, in spite of the problem of commitment that this practice introduces. In my model, instead, when issuing refresher grants or repricing is optimal, the principal commits to it ex–ante and only in the appropriate nodes of the game.[13]

A third related paper is Edmans et al. (2012). The authors use for their analysis a particular hidden information repeated moral hazard setup that allows them to provide an elegant closed form solution for a complete characterization of the optimal contract.[14] In their framework, the optimal contract can be implemented using a scheme that establishes an escrow account for the CEO in the initial period. Then, the contract dictates what proportion of the account becomes available for consumption over time. Moreover, the proportion of the balance in the account that must be invested in stock of the firm needs to be "rebalanced" over time, in response to changes in the value of the firm given new signal realizations, in order to keep implementing the optimal incentives for

[12] One main difference is that options granted outside of a shareholder pre–aproved long term option plan are not tax deductible under current tax laws (see the Omnibus Budget Reconciliation Act resolution 162(m) of 1992); hence, repricing may be an attractive alternative for companies that have exhausted the options available in their option plans.

[13] It will be clear that the equilibrium of my model is not robust to allowing for ex–post Pareto improving renegotiations. Once the effort is done, it is optimal to renegotiate to uncontingent payments in the consumption stage. However, this is anticipated by the agent ex–ante, and implementing high effort is no longer feasible.

[14] In Edmans et al (2012), each period's signal is equal to the sum of a hidden action and a stochastic noise variable. In a departure from the standard hidden *action* moral hazard problem in Grossman and Hart (1983), they assume that the agent observes the noise realization before he chooses his effort (a hidden *information* moral hazard model). This timing simplifies the problem of finding the optimal contract, and it is also necessary for the implementation with a scheme that uses only stock and cash (see their discussion on page 1630 of their article).

the CEO. Despite the differences across their setup and mine, the rebalancing of their incentive account and the refresher grants in my model respond to the same force: the need to adjust the sensitivity of compensation packages that are contingent on stock prices when new information is revealed about the value of the firm that affects those prices differently than it affects the optimal contract.

The paper is organized as follows. Section 2 presents the model. The equilibrium is analyzed in section 3, with the results for unconstrained contracts in section 3.2 and those for real life compensation instruments in section 3.3. A generalization of the model is discussed in section 4, and section 5 concludes. All proofs are relegated to the Appendix.

2 The model

I model the moral hazard problem that arises between the CEO and the owners of a firm due to the unobservability of the CEO's effort. I assume the CEO (agent) is risk averse with $u(c) = \ln(c)$. The firm is owned by well diversified shareholders that coordinate perfectly to act as the unique risk–neutral principal of the agent. I also assume that there is a competitive stock market that prices the stock of the firm according to its expected per–period output. I assume that the owners of the firm can commit to implement a compensation contract.

The contract lasts for two periods, $t = \{1, 2\}$. The output of the firm can take two values each period, $y_L = 0$, and $y_H = 1$. The agent affects the probability distribution over output with his effort, which can take two values: e_L, with disutility of effort $v(e_L) = 0$, or e_H with $v(e_H) = e$. This effort is done only once, at the beginning of period 1, but it affects the distribution of output both at period 1 and period 2. Hence, the effect of effort is persistent in time. The agent receives his payment only at the end of the second period. The first period represents an interim stage at which new information is revealed (the first period output realization is observed), but no consumption takes place in it. For simplicity, I also assume that there is no discounting between periods 1 and 2.

The distribution over output is also affected by another parameter: a state that determines the effectiveness of the effort of the CEO, denoted $\theta \in \{A, B\}$. The true state is unknown by the agent, the principal, and the stock market, and all players attach a prior probability of q_0 to $\theta = A$. The probability of observing a high output contingent on an effort level and a realization of the state is as follows, for every t:

$\Pr(y_t = y_H \mid e, \theta)$	(q_0) A	$(1 - q_0)$ B
e_H	π_A	π_B
e_L	$\widehat{\pi}_A$	$\widehat{\pi}_B$

(1)

For my leading example firm, I assume $\pi_A = \pi$, $\widehat{\pi}_A = \widehat{\pi}$, and $\pi_B = \widehat{\pi}_B = 1$. In such a firm, when effort is not effective, the firm always produces high output. This is a simplifying assumption that I will relax in section 4, when I will consider firms for which output is always low in state

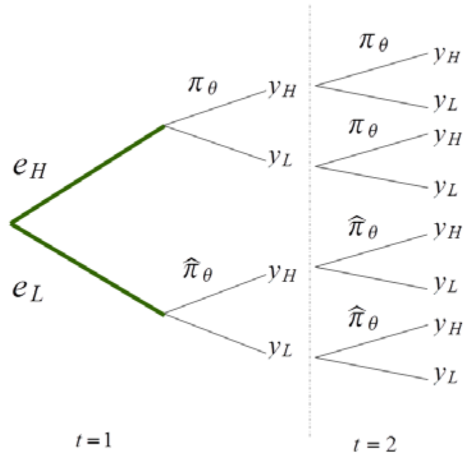

Figure 3: Timing and probabilities of output realizations. The state is represented by $\theta = \{A, B\}$.

B ($\pi_B = \widehat{\pi}_B = 0$), and firms for which effort is also effective in state B, but it implements different probabilities than in state A. To distinguish this type of firm in matrix 1 from other types introduced later, I will refer to it as a type $\mathcal{H}$ firm.

All probabilities are common knowledge. I assume that the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H in $\theta = A$, that is, $1 > \pi > \widehat{\pi} > 0$. The timing and the stochastic structure are depicted in Figure 3. The above assumptions on the probabilities imply that, at time 0, all the nodes of the tree have positive probability of being reached under both levels of effort.

Matrix 1 constitutes a very stylized model of a firm's technology. However, moral hazard and learning are present in this technology, complicating the analysis of compensation that is the objective of this paper.

2.1 Learning about the state θ

Each period, after observing the output realization, the principal, the agent, and the stock market update their prior about the quality of the match. The updating is done using Bayes' rule, for a given e choice. I denote the posteriors in the first period when the agent chooses e_H as q_i, and those in the second period as q_{ij}, for $i = L, H$, $j = L, H$. Similarly, the posteriors when the agent chooses e_L are denoted $\widehat{q}_i$ and $\widehat{q}_{ij}$. To simplify the exposition, I introduce the following notation:

$$\begin{aligned}
\pi_0 &= q_0 \pi_A + (1 - q_0)\, \pi_B \\
\pi_i &= q_i \pi_A + (1 - q_i)\, \pi_B, \quad i = L, H, \\
\pi_{ij} &= q_{ij} \pi_A + (1 - q_{ij})\, \pi_B, \quad i = L, H,\ j = L, H,
\end{aligned}$$

which denote the probability attached by all players to observing a high output realization in the first period (π_0), and in the second period, contingent on the first period realization (π_i), and, in the third period, contingent on both realizations (π_{ij}). Similarly, for e_L, the corresponding probabilities are denoted by $\widehat{\pi}_0$, $\widehat{\pi}_i, \widehat{\pi}_{ij}$.

2.2 Valuation of the firm by outside investors

I assume that there are a large number of investors in the economy who are willing to buy the stock of the firm. Investors value the stock of the firm as a claim to the expected output of the firm.[15] I also assume a large number of shareholders (sellers of the stock), so no individual deviation affects the equilibrium price. Competition implies a price equal to the expected output. Investors and shareholders update their beliefs about θ, for a given effort e. Hence, the market price for the stock varies as output realizations become available.

In order to simplify my analysis of the compensation problem, I normalize the price to the expected per–period value of the firm:[16]

$$p\left(y^t, e\right) \equiv E_t\left[y_t \mid e\right],$$

where $E_t[\cdot]$ denotes the expectation taken with all information known at t, which is summarized by the posteriors based on the history of realizations, y^t. I introduce the following notation for prices. The price of the stock corresponds to the expected value of the firm given the history of realizations and a given effort choice:

$$p_0 \equiv p\left(\emptyset, e_H\right) = \pi_0, \tag{2}$$

$$p_i \equiv p\left(y_i, e_H\right) = \pi_i \quad i = L, H, \tag{3}$$

$$p_{ij} \equiv p\left(y_i, y_j, e_H\right) = \pi_{ij} \quad i = L, H,\ j = L, H, \tag{4}$$

under high effort. Similarly, under low effort, $\widehat{p}_0, \widehat{p}_{ij}$, and $\widehat{p}_{ij}$ will be equal to $\widehat{\pi}_0, \widehat{\pi}_i$, and $\widehat{\pi}_{ij}$, correspondingly.

2.3 Compensation Instruments

In this section I define the compensation instruments available to the firm. I allow the compensation package to include the following elements: an annual wage, a bonus plan, ad–hoc payments, and long–term performance–based plans that include both stock and option grants. With these elements, I try to capture the most important features of real–life compensation practices.[17] In 2010,

[15] By appliying this pricing rule in the second period I implicitly assume that the technology of the firm lasts for at least 3 periods (i.e., high otput happens with the same probability in $t = 3$ as in $t = 1$ or 2). The commitment to the contract, however, lasts only until $t = 2$, so the third period realizations cannot be used for compensation. This simplification allows me to derive the analytical results in the paper.

[16] If instead prices where equal to the net present value of output, the price in the first period would represent a claim to two output realizations, while the price in the second period would be a claim only to one output. This would imply a difference in the level of prices that is irrelevant for the economic problem of interest in this paper, but it complicates the algebra.

[17] See Murphy (1999) for a detailed description of compensation instruments based on compensation surveys. See Hall and Liebman (1998) for details based on hand–picked data from proxy statements. See Clementi and Cooley (2009) for a recent and careful description of the main facts related to the level and structure of compensation of the executives of the largest public U.S. firms in the last two decades using Execucomp data.

data in Execucomp for the CEOs of the 1,500 largest public companies in the United States shows that the average pay was \$4,371,060, with a minimum of \$200,000 and a maximum of \$25,761,432. The median of the highly skewed distribution of pay was of \$3,022,000.[18] Of this total pay, the salary represented an average of 25% (or a median of 19%), the bonus and incentive program represented an average of 25% (or a median of 23%), stock grants 28% (median of 25%), option grants an average of 19% (median of 13%), and perks and other compensation an average of 3% (median of 1%).

I now present a brief description of each instrument and how it is captured in the model.

Base Salaries In real life, salaries for CEOs are normally negotiated at the time of signing a contract, based on industry benchmarks. The negotiation usually includes a pre–specified annual increase for the duration of the contract, independently of performance. In the model, the salary is a constant payment given in period 2. I denote the salary as W.[19]

Bonus plans In real life, most companies offer bonus plans paid annually based on the firm's performance as measured by accounting results. They usually specify a performance target, together with a minimum and maximum limit for bonuses and the sensitivity of the bonus to the performance measure. These performance measures consist mainly of objective measures such as net income, revenue, pre–tax income, or other accounting figures. Typically, about 25% of the total measures used in the evaluation are labeled as "individual performance" measures, which are subjective evaluations. In the model, I summarize these characteristics by making the bonus plan depend only on accumulated annual output, $(y_1 + y_2)$, as follows: $B^c(y_1, y_2) = \min\{b^c, b^c(y_1 + y_2)\}$. This mimics the structure of bonus programs in real life, where a "pool" available for bonuses determines a "cap" (b^c, in this case) for annual payments.[20]

Long–term incentive plans In real life, compensation plans include long–term payments mainly in the form of (i) stock of the company and (ii) options to buy stock at a pre–determined price (the "exercise" price or "strike" price.) Both typically come with selling restrictions: They cannot be traded before their "vesting" time. Also, the manager cannot take these grants with him if he leaves the company, and he is not allowed to hedge against the risk in his compensation package. In the model, I assume that all grants vest in period 2, and they are exercised immediately by the CEO. Consistently with real–life practices, I assume an exercise price equal to the market price of the stock at the time of granting. There is evidence that most firms use multiyear stock and option plans to determine the value of annual grants. They also occasionally use out–of–cycle or larger than average grants, often

[18] This calculation excludes CEOs that at some point in their tenure with their firms owned more than 3% of the total stock of their company; in picking this threshold I follow Clementi and Cooley (2009), who argue that a high ownership is not consistent with a moral hazard problem.

[19] Pension payments and life insurance premiums are usually offered to executives as part of their compensation; because their relative importance on the overall compensation is small, I do not explicitly include them in the model. However, given their noncontingent nature, one could assume that they are included in the variable W in the model.

[20] Results for an alternative linear bonus program, defined as $B^l(y_1, y_2) = b^l(y_1 + y_2)$, are similar and are discussed in section 4.

called "refresher" grants; I model these as grants given in the interim period.[21] I simplify the vesting restrictions and expirations and assume that all stock and option grants can and must be sold or exercised at time 2. I denote the different grants as follows:

- r_0: restricted stock issued in period 0
- r_i, for $i = L, H$: refresher restricted stock issued in period 1, contingent on realization i being observed in period 1
- s_0 : stock option grant issued at time 0, with exercise price p_0
- s_i, for $i = L, H$: refresher stock option grant issued at time 1, contingent on realization i being observed in period 1, with exercise price p_i

Ad–hoc payments This category includes compensation that is not granted to the executive on a regular basis, and for which the amounts received by the executive are not a previously set function of accounting measures or stock price realizations. Examples are personal benefits ("perks"), subsidized loans, signing bonuses, severance payments, or discounted share purchases. I will denote these ad–hoc payments as k_{ij}, for $i, j = L, H$, to make clear that these payments may be fully contingent on the history of realizations of output.

For incentive purposes, it is important that the CEO understands and anticipates what the compensation contract implies for his consumption in each possible state of the world, including both the return of standard compensation instruments like bonus plans or grants, as well as ad–hoc payments. In the model, I capture this feature by assuming commitment to the contract. However, I will distinguish between compensation packages that include standard, transparent compensation instruments that are a function of accounting results (like a bonus plan) or stock price realizations (like stock or option grants), versus ad–hoc transfers. I make this distinction formally in the next subsection.

2.4 Compensation Packages and Consumption for the CEO

I denote a compensation package as a vector $P = \left(W, b^c, r_0, s_0, \{r_i\}_{i=L,H}, \{s_i\}_{i=L,H}, \{k_{ij}\}_{i,j=H,L}\right)$. Let $\mathbb{P} \subset \mathbb{R}_+^{10}$ be the set of all possible compensation packages. Based on whether ad–hoc payments are used, and on whether the package includes more complicated instruments like options, I define the following strict subsets of $\mathbb{P}$:

$$\begin{aligned} \mathbb{S} &= \{P \in \mathbb{P} \text{ such that } s_0 = r_i = s_i = k_{ij} = 0 \ \forall i, j\}, \\ \mathbb{C} &= \{P \in \mathbb{P} \backslash \mathbb{S} \text{ such that } k_{ij} = 0 \ \forall i, j\}. \end{aligned}$$

[21] Hall (1999) reports that most of the firms in a sample of large publicly traded U.S. firms use option plans that imply either a fixed number of shares to be granted every year for the duration of the plan (about 40% of the sample firms), or a fixed value of grants (where the number is adjusted according to stock price changes). He classifies a firm under one of these plans if he observes the same number or value for two consecutive years. Hall (1999) defines "refresher" grants as "out–of–cycle" or larger than usual grants. See also Hall and Knox (2004), which I discuss later in this paper in relation to my results, for evidence on refresher grants.

Definition 1 A compensation package P is classified as:

- **Transparent** if it does not include any ad–hoc payments, i.e. $P \in \mathbb{C} \cup \mathbb{S}$,
  - **Simple**: if it is transparent and it includes only a wage, bonus scheme, and restricted stock granted at time 0, i.e., $P \in \mathbb{S}$.
  - **Complex**: if it is transparent and it includes at least one option grant or a refresher stock grant but no ad–hoc payments, i.e., $P \in \mathbb{C}$.
- **Non-transparent**: if it includes at least one ad–hoc payment, which has a value that is not a set function of output realizations or stock prices, i.e., $P \notin \mathbb{C} \cup \mathbb{S}$.

I assume that the principal can force the agent to sell all stock and options in period 2 when they are profitable and consume all income generated from it. Hence, given any $P \in \mathbb{P}$, I can calculate the implied consumption of the agent. I introduce the following notation to denote this consumption as a function of the compensation package: $\mathcal{C}(P) = \{c_{ij}\}_{i,j=L,H}$. This function takes the following form:

$$\begin{aligned}
c_{HH} &= W + b^c + r_0 p_{HH} + s_0 (p_{HH} - p_0) + r_H p_{HH} + s_H (p_{HH} - p_H) + k_{HH} \\
c_{HL} &= W + b^c + r_0 p_{HL} + \max\{s_0 (p_{HL} - p_0), 0\} + r_H p_{HL} + k_{HL} \\
c_{LH} &= W + b^c + r_0 p_{LH} + \max\{s_0 (p_{HL} - p_0), 0\} + r_L p_{LH} + s_L (p_{LH} - p_L) + k_{LH} \\
c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} + k_{LL}.
\end{aligned} \tag{5}$$

It is important to note that the function $\mathcal{C} : \mathbb{P} \to \mathbb{R}^4_+$ is not injective, that is, different compensation packages may imply the same contingent consumption vector. Using the function $\mathcal{C}$, we can calculate the expected utility of the agent. For a given compensation package $P \in \mathbb{P}$ and high effort, this expected utility is:

$$\begin{aligned}
U(P, e_H) &= \pi_0 [\pi_H \ln(c_{HH}) + (1 - \pi_H) \ln(c_{HL})] \\
&\quad + (1 - \pi_0) [\pi_L \ln(c_{LH}) + (1 - \pi_L) \ln(c_{LL})] - e.
\end{aligned}$$

Finally, the cost to the principal of a contract P that implements e_H is

$$\begin{aligned}
K(P, e_H) &= \pi_0 [\pi_H c_{HH} + (1 - \pi_H) c_{HL}] \\
&\quad + (1 - \pi_0) [\pi_L c_{LH} + (1 - \pi_L) c_{LL}].
\end{aligned}$$

The expected utility under low effort, $U(P, e_L)$, and the cost of implementing low effort, $K(P, e_L)$, are constructed in a similar manner, changing the probabilities to those corresponding to low effort.

2.5 Incentive problem

With the compensation packages and the consumption function in hand, we are now ready to write the optimization problem of the principal. I assume throughout that parameters are such that it is always profitable to implement e_H. Hence, the optimal compensation package P^* is the solution to the following cost minimization problem, where $\underline{U}$ represents the outside utility the agent would obtain if he were not to participate in the contract:

$$V(P) = \min_{P \in \mathbb{P}} \quad K(P, e_H) \quad \text{s.to} \tag{\textbf{P1}}$$

$$\underline{U} \leq U(P, e_H) \tag{PC}$$

$$U(P, e_H) \geq U(P, e_L) \tag{IC}$$

Problem P1 is difficult to solve in general, due to the large number of non–negativity constraints that the domain for P implies. Moreover, the function $\mathcal{C}(P)$ depends on the equilibrium stock prices, which in turn depend on the choice of effort. I propose, instead, to solve a simplified problem in which the principal chooses directly a tuple $C = \{c_{ij}\}_{i,j=L,H}$ of transfers contingent on the history of output realizations, as follows:

$$V(C) = \min_{C} \quad K(C, e_H) \quad \text{s.to} \tag{\textbf{PS}}$$

$$\underline{U} \leq U(C, e_H) \tag{PC'}$$

$$U(C, e_H) \geq U(C, e_L) \tag{IC'}$$

$$c_{LL}, c_{LH}, c_{HL}, c_{HH} \geq 0. \tag{NNC'}$$

I denote the solution to PS as $C^* \equiv \left\{c^*_{ij}\right\}_{i,j=L,H}$. The standard arguments valid for a static moral hazard problem (see Grossman and Hart, 1983) justify that both the PC and the IC constraints bind in the optimum. Note that the agent has logarithmic utility, so the non–negativity constraints (which are now in terms of consumption levels) will never bind. Also, with a simple change of choice variables to utility levels, the objective function is linear and the constraint set is compact and convex, so the solution to PS exists and is unique.

Lemma 1 *Any solution P^* to problem P1 implements the same consumption for the agent as the solution C^* to the simplified problem PS.*

The proof is obvious so it is omitted. It is easy to see that the set of available compensation instruments $\mathbb{P}$ is rich enough to implement any (positive) transfer scheme contingent on the history of output realizations, i.e., any value for the tuple $\{c_{ij}\}_{i,j=L,H}$. In particular, setting $k_{ij} = c^*_{ij}$ for all i, j is always a sufficient implementation. This result implies that I can study the problem of choosing the instruments separately from the determination of contingent consumption in the optimal contract. However, since the function $\mathcal{C}(P)$ is not invertible there might be several compensation packages that solve problem $P1$ and satisfy $\mathcal{C}(P) = C^*$

3 Equilibrium

Recall from section 2.2 that individual deviations of the shareholders and the investors do not affect the stock prices in the equilibrium of the stock market. This implies that there are only two pricing rules that may appear in any equilibrium: one for any contract that implements e_L and one for any contract that implements e_H. Moreover, when considering a deviation, the CEO realizes that he can only affect the probability distribution over prices but not the prices themselves.

An equilibrium of the above game between the principal, the agent, and the stock market is defined next.

Definition 2 A Perfect Bayesian Equilibrium of this game in which effort e_H is implemented consists of a compensation contract P^* and stock prices such that

a) $\mathcal{C}(P^*) = C^*$, where $C^* = \arg\min_{\{C\}} K(C, e_H)$

b) The utility of the agent choosing e_H is equal to that of choosing e_L and is as large as his outside utility $\underline{U}$

c) Market prices and the beliefs of the stock market participants about θ are consistent with the agent choosing e_H, as defined by $p_0, p_L, p_{LL}, p_{LH}, p_{HL}, p_{HH}$ in Equations (2) -(4)

d) Beliefs about θ are updated according to Bayes' rule

Since the probability of observing any history is positive under the equilibrium level of effort, Bayesian updating provides consistent beliefs, and no refinement is necessary.

In the next subsections, I describe the properties of the equilibrium.

3.1 Equilibrium Stock prices

All stock traders anticipate that, in equilibrium, the agent chooses the recommended level of effort, e_H. Hence, they update their beliefs using the probabilities in the above matrix corresponding to e_H. The equilibrium price of the stock corresponds to the expected output of the firm given the history of realizations, given by p_i and p_{ij} in Equations (2) -(4). For the rest of the analysis in the paper, it will be useful to keep in mind the following property of stock prices:

Lemma 2 *Stock prices are monotonic in the period's output:* $p_L \leq p_H$, *and* $p_{LL} \leq p_{LH} = p_{HL} \leq p_{HH}$.

In particular for a type $\mathcal{H}$ firm, for any histories containing at least one y_L, the updated beliefs put probability one on $\theta = A$, i.e., we have that $q_{ij}^{\mathcal{H}} = 1$ if i or j equals L. If the observed history does not contain any y_L, instead, $\theta = B$ has still positive probability. This is the case for histories y_H and (y_H, y_H). That is, using Bayes' rule,

$$\begin{aligned} q_H &= q_0 \frac{\pi}{q_0 \pi + 1 - q_0}, \\ q_{HH} &= q_0 \frac{\pi^2}{q_0 \pi^2 + 1 - q_0}, \\ q_L &= q_{LH} = q_{LL} = 1. \end{aligned}$$

A direct implication of this learning is that the stock prices take the simple form:

$$\begin{aligned} p_0 &= q_0 \pi + 1 - q_0, \\ p_H &= q_H \pi + 1 - q_H, \\ p_{HH} &= q_{HH} \pi + 1 - q_{HH}, \\ p_L &= p_{HL} = p_{LL} = \pi. \end{aligned} \tag{6}$$

3.2 Equilibrium Consumption

Problem **PS** is a particular example of a static moral hazard problem, with i.i.d. output revealed over time and exogenous uncertainty about the probability distribution implemented by each effort level. The characterization of the optimal contract with unrestricted instruments follows easily from the standard first order conditions of problem **PS**. Define the likelihood ratio (LR) of a history of realizations as the ratio of the expected probabilities of that history under low and high effort:

$$\begin{aligned} LR_{HH} &= \frac{\widehat{\pi}_0}{\pi_0} \frac{\widehat{\pi}_H}{\pi_H} = \frac{q_0 \widehat{\pi}^2 + 1 - q_0}{q_0 \pi^2 + 1 - q_0}, \\ LR_{HL} &= \frac{\widehat{\pi}_0}{\pi_0} \frac{1 - \widehat{\pi}_H}{1 - \pi_H} = \frac{(1 - \widehat{\pi}) \widehat{\pi}}{(1 - \pi) \pi}, \\ LR_{LH} &= \frac{1 - \widehat{\pi}_0}{1 - \pi_0} \frac{\widehat{\pi}_L}{\pi_L} = \frac{(1 - \widehat{\pi}) \widehat{\pi}}{(1 - \pi) \pi}, \\ LR_{LL} &= \frac{1 - \widehat{\pi}_0}{1 - \pi_0} \frac{1 - \widehat{\pi}_L}{1 - \pi_L} = \frac{(1 - \widehat{\pi})^2}{(1 - \pi)^2}. \end{aligned} \tag{7}$$

Proposition 1 *Consumption levels in the optimal contract C^* are ranked by likelihood ratios:*

$$c_m^* > c_n^* \Leftrightarrow LR_m < LR_n \quad , \text{ for } n, m \in \{LL, LH, HL, HH\}.$$

Moreover, consumption is linear in the LR:

$$c_m^* = \lambda + \mu (1 - LR_m), \tag{8}$$

where λ is the multiplier of the constraint PC and μ that of the constraint IC.

It is worth noting that, if θ were A for sure, the above characterization would always imply the same ranking for consumptions for all combinations of parameters, as the next proposition states. Define $\Delta_L \equiv c_{LH}^* - c_{LL}^*$ and $\Delta_H \equiv c_{HH}^* - c_{HL}^*$.

Proposition 2 *In the absence of learning about θ (certainty case with $\theta = A$), the optimal consumption is monotonic in cumulative output and it satisfies:*

$$0 < \Delta_H < \Delta_L.$$

With uncertainty about θ, however, the posterior evolves differently under e_L than e_H, changing the weight of each probability in the numerator and denominator of the LR. As stated in the next proposition, this can create non–monotonicities: For a type $\mathcal{H}$ firm, it may be the case that, when the first period output has been y_H, the agent's wage may be higher if we observe y_L in the second period than if we observe y_H.

Proposition 3 *When the firm is of type $\mathcal{H}$, consumption spread always satisfies $\Delta_H < \Delta_L$. Also, non–monotonicities never arise following a low output in the first period, i.e., $\Delta_L > 0$ always. Moreover, we have*

(i) *whenever $\pi + \widehat{\pi} \geq 1$, for all $q_0 \in (0,1)$, $\Delta_H > 0$,*

(ii) *whenever $\pi + \widehat{\pi} < 1$,*

$$\begin{aligned} \text{for } q_0 &\in (0, q^*), \ \Delta_H < 0, \\ \text{for } q_0 &\in (q^*, 1), \ \Delta_H > 0, \end{aligned}$$

where $q^* = \frac{1-\pi-\hat{\pi}}{1-\pi-\hat{\pi}+\pi\hat{\pi}}$.

The fact that non-monotonicities only arise in Δ_H, and only when $\pi + \widehat{\pi} < 1$, is related to the interaction of the informativeness of the signal (y_L, y_H) (or, equivalently, (y_H, y_L)) and the learning about the true state. On one hand, $\pi + \widehat{\pi} < 1$ implies that LR_{LH} is less than one, and hence the optimal contract seeks to reward the agent when observing LH, as well as when observing HH. Punishments are reserved for LL. However, observing LH reveals perfectly that the true state is A, making y_H a more informative signal about effort than if there were still positive probability on state B (which is the case when we observe HH). This tends to make c_{LH} large, but not c_{HH}, making Δ_L large and Δ_H small. For low enough q_0, the relative informativeness of LH and HH may be reversed and we may get $\Delta_H < 0$.

Figure 4 presents graphically the analytical characterization in Proposition 3 of the parameters that imply a nonmonotonicity. For this, we assume that, for each sufficient combination of π and $\widehat{\pi}$, there is a mass one of firms whose prior q_0 is distributed uniformly between 0 and 1. The vertical axis, then, represents the proportion of firms for which nonmonotonicity is present in the optimal contract. We see that, when π is high enough, or the difference $\pi - \widehat{\pi}$ is large enough (both cases that lead to $\pi + \widehat{\pi} > 1$), consumption is monotonic. For the combinations that imply the possibility of nonmonotonicities, with $\pi + \widehat{\pi} < 1$, consumption will be nonmonotonic for the firms with the smaller priors, $q_0 < q^*$ as defined in Proposition 3. Hence, figure 4 is, in practice, a graph of the threshold $q^* (\pi, \hat{\pi})$.

With the properties of the optimal contract in hand, we now turn to the implications for compensation packages.

3.3 Equilibrium compensation packages

The analysis of the properties of equilibrium consumption in the previous section was based on the solution to problem **PS**, with contingent consumption transfers C^*. In this section, I use the $\mathcal{C}(C^*)$

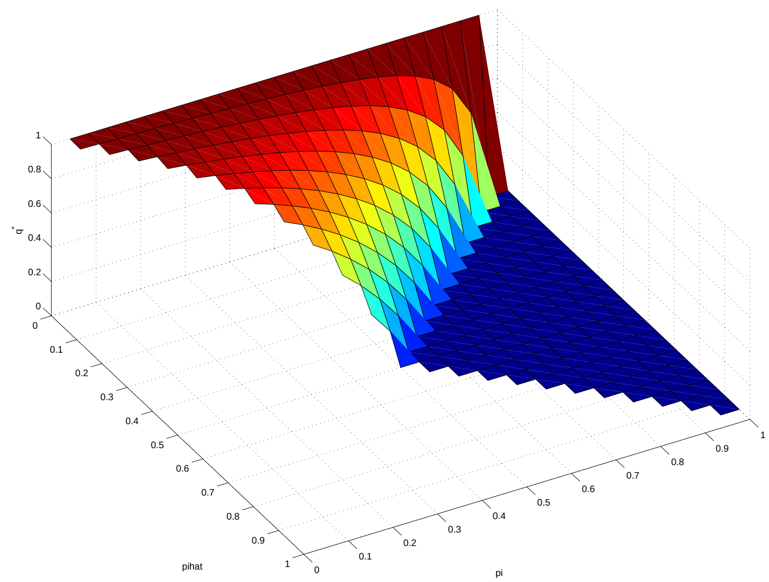

Figure 4: Type $\mathcal{H}$ firm. Probability that non–monotonicities arise in the optimal contract ($c_{HH} < c_{LH}$).This happens whenever $q_0 \leq q^*$.

mapping, together with the properties of C^*, to analyze the characteristics of the solution to the original problem **P1** in terms of compensation packages P in $\mathbb{P}$.

The first thing to note is that given the richness of the elements of $\mathbb{P}$, the optimal contract C^* characterized in the previous subsection is always sufficient in a trivial way: Because of the availability of ad–hoc payments, the firm can simply set $k_{ij} = c^*_{ij}$ for all ij pairs. However, there may be other combinations of compensation instruments that implement a given optimal contract. To solve the indeterminacy of the compensation package, I make the following assumption, which makes use of the classification of compensation packages presented in page 13:

Assumption 1 The principal, when presented with several choices to implement a given contingent consumption scheme, prefers a simple scheme to a complex one, and prefers not to use ad–hoc payments.

Implicit behind this assumption is the fact that it may be more costly for shareholders to use ad–hoc payments or complex compensation instruments. The costs may include communication to investors of compensation practices or tax deductions that cannot be taken advantage of. In the rest of this section, I ask the following questions: What types of firms are not able to avoid using ad–hoc payments, and which are? Which can do with just a simple scheme? To answer these questions, I use the following strategy. First, I spell out $\mathcal{C}(P)$ under the restrictions implied by a simple and a complex scheme. Then, I analyze the system of equations resulting from equating $\mathcal{C}(P) = C^*$.

Definition 3 Consider a firm defined by a probability structure of the form of matrix (1). A compensation scheme P is **sufficient** if, for the C^* corresponding to the parameter values that describe the firm, the system of equations resulting from equating $\mathcal{C}(P) = C^*$ has a solution.

Note that although restricting to schemes in the subsets $\mathbb{S}$ and $\mathbb{C}$ simplifies the system $\mathcal{C}(P) = C^*$, it is difficult to characterize the solution in general, since all packages P need to have non–negative elements. In what follows, I present a partial characterization of the choice of compensation packages between simple, complex, or nontransparent. I complement my analysis with a complete numerical characterization of this choice.

The conditions in the propositions in this section inform us about the restrictions that using standard compensation instruments imposes on the sensitivity of the consumption of the agent to changes in stock prices. This sensitivity is sort of a reduced form for the composition of the sensitivity of consumption to signals (i.e., output realizations), which is dictated in the optimal contract by the likelihood ratios, and the sensitivity of prices to signals, which is in turn dictated by the pricing rule of outside investors. The results that I present next show that a limited set of compensation instruments like $\mathbb{S}$ puts severe restrictions on the relationship of these two sensitivities, but still there are many parameter combinations for which these limited instruments are rich enough to implement the optimal contract. Complex schemes, on the other hand, are always sufficient when consumption is monotonic. Also, they do not necessarily include more than three instruments, although one of them will need to be a refresher option grant.

3.3.1 When are firms able to avoid ad–hoc payments? Sufficiency of transparent schemes

By simple inspection of the system $C^* = \mathcal{C}(P)$, we can see that payments to the agent coming from a compensation package that does not include ad–hoc payments are necessarily monotonic in prices, and hence monotonic in output, since prices are themselves monotonic in output. As the following proposition describes, monotonicity of the optimal consumption is both a necessary and a sufficient condition for a complex scheme to be sufficient.

Proposition 4 *A complex scheme is sufficient for a type $\mathcal{H}$ firm (with a capped bonus) if and only if $\Delta_H \geq 0$, i.e., if and only if $q_0 \leq q^* = \frac{1-\pi-\hat{\pi}}{1-\pi-\hat{\pi}+\pi\hat{\pi}}$.*

The proof of this proposition shows that the system $C^* = \mathcal{C}(P)$ for a complex scheme is undetermined, i.e., if it has a solution, it has an infinite number of them. However, only solutions that satisfy the non–negativity constraints on all the instruments constitute sufficient complex schemes. In the proof of the sufficiency it is shown that whenever $\Delta_H \geq 0$, the following scheme is always sufficient:

$$\begin{aligned} W &= c_{LL} \\ b &= \Delta_L \\ s_H &= \frac{\Delta_H}{p_{HH} - p_H} \\ s_0 &= r_0 = r_L = r_H = s_L = 0. \end{aligned} \tag{9}$$

This scheme includes only three instruments, but one of them is a "refresher" option grant that is promised to the agent only if the first period output is high. Hence, the exercise price of this option grant will be p_H, and it will only pay off in the final node HH.

A different way of reading proposition 4 is that, for a firm of type $\mathcal{H}$, compensation schemes must be nontransparent if and only if non–monotonicities arise $(\Delta_H < 0)$. Figure 4 in page 18 represented the population of firms for which this was the case. The model implies that only for these firms are ad–hoc payments unavoidable.

3.3.2 When are firms able to avoid options and refresher grants? Sufficiency of simple schemes

In this section, I provide necessary and sufficient conditions for the sufficiency of a simple compensation package, which does not include options or "refresher" grants given in period one after the interim output is realized. One caveat is that the conditions will not be, in general, in terms of the primitive parameters of the firm; to gain insight on the sufficiency in terms of primitives, I complement the analytical derivations with numerical characterizations.

Proposition 5 *A simple scheme is sufficient if and only if*

a) $\frac{c^*_{LL}}{\Delta_H} \geq \frac{p_{LL}}{p_{HH}-p_{HL}} = \frac{\pi}{(1-q_{HH})(1-\pi)}$, *and*

b) $\Delta_H \geq 0$, or, equivalently, $q_0 \leq q^* = \frac{1-\pi-\hat{\pi}}{1-\pi-\hat{\pi}+\pi\hat{\pi}}$ *(see Proposition 3).*

The proposition can be otherwise stated as a simple scheme being sufficient whenever the solution $C^* = \mathcal{C}(P)$ for a simple scheme satisfies the non–negativity constraint for the three instruments:

$$\begin{aligned} W &= c^*_{LL} - \pi \frac{\Delta_H}{(1-\pi)\left(1-q^{\mathcal{H}}_{HH}\right)} \geq 0, \\ b^c &= \Delta_L \geq 0, \\ r_0 &= \frac{\Delta_H}{(1-\pi)\left(1-q^{\mathcal{H}}_{HH}\right)} \geq 0. \end{aligned}$$

The intuition behind the form of this solution is conveyed next. For a firm of type $\mathcal{H}$, any number of low outputs is perfectly informative about the state. Hence, there is no variation in prices in the lower range, implying that the spread $c_{LH} - c_{LL}$ needs to be implemented with the bonus payout: $b = \Delta_L$. Setting the bonus b to satisfy this constraint is always sufficient, since $\Delta_L > 0$ always for a firm of type $\mathcal{H}$. Since we also have that $\Delta_H < \Delta_L$ always, the bonus plan needs to be capped,

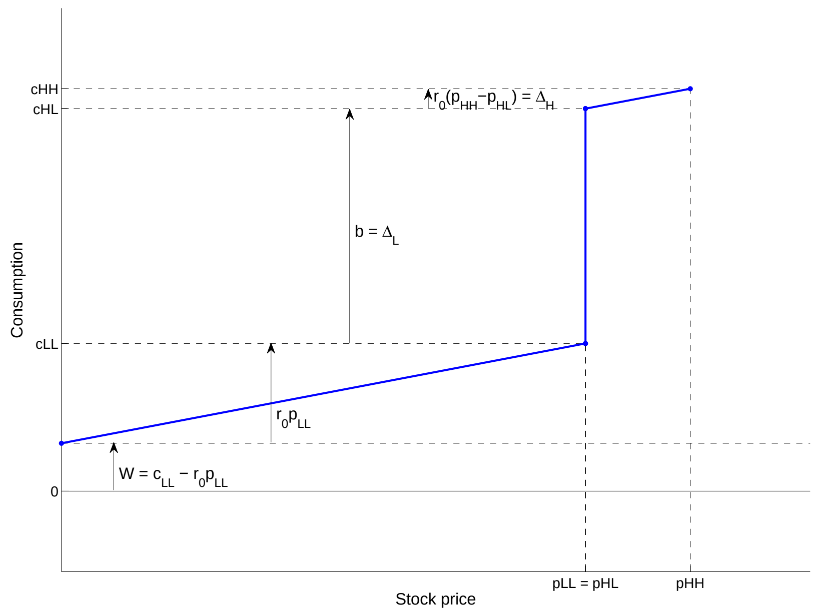

Figure 5: For the firm in Example 1, a simple scheme is feasible.

rather than linear in cumulative output. Hence, the spread $c_{HH} - c_{HL}$ needs to be implemented with restricted stock:

$$\Delta_H = r_0 \left(p_{HH} - p_{LH}\right).$$

This is feasible only if the optimal consumption is monotonic, so that we have $\Delta_H > 0$ (condition b) in proposition 5). Finally, the level of c_{LL} needs to be implementable, given the restricted stock r_0, with a positive wage:

$$W = c_{LL} - r_0 p_{LL}.$$

This is feasible whenever condition a) in in proposition 5 is satisfied. Example 1 presents a firm for which a simple scheme is sufficient.

Example 1 Consider a type $\mathcal{H}$ firm described by the following parameters: $\pi = .7$, $\hat{\pi} = .28$, and $q_0 = .3$. The agent has an outside utility of $\bar{U} = 5$, and an effort disutility of $e = .3$. The optimal contract is: $c_{LL} = 79.5$, $c_{HL} = 205.9$, and $c_{HH} = 216.6$. The equilibrium stock prices are: $p_0 = .79$, $p_L = p_{LL} = p_{LH} = .7$, $p_H = .81$, and $p_{HH} = .84$. For this firm, a simple scheme is sufficient, and it takes the following values: $W = 25.8$, $b^c = 126.3$, and $r_0 = 76.8$. Figure 5 presents this scheme graphically.

Unfortunately, condition a) in Proposition 5 depends in a nontrivial way on the primitives of the model (through $\frac{c^*_{LL}}{\Delta_H}$), and it is not possible to provide an analytical characterization of the combination of $\bar{U}, e, q_0, \pi$ and $\hat{\pi}$ for which it is satisfied. As lemma 3 in the appendix states,

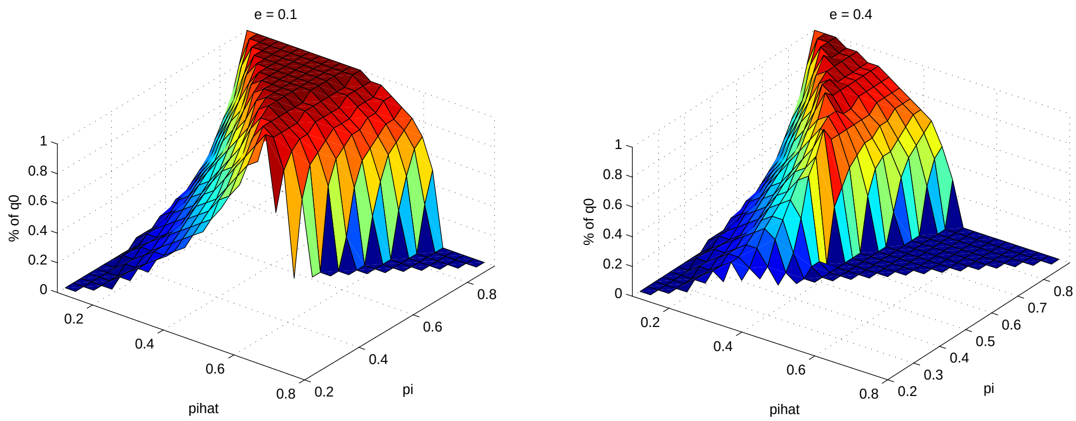

Figure 6: Fraction of type $\mathcal{H}$ firm firms for which a simple scheme is sufficient, for $e = .1$ (left panel) and $e = .4$ (right panel). There is sufficiency for a smaller set of firms when effort disutility is higher.

however, condition a) is independent of $\bar{U}$.[22] This result provides us with a simpler way of checking the condition numerically. Figure 6 plots in the left panel the probability that both condition a) and b) are satisfied for a given π and $\hat{\pi}$ combination, for $e = .1$, and assuming a uniform distribution over an evenly spaced grid of q_0. The values that make it most likely are combinations of high values of π with intermediate values of $\hat{\pi}$. Whenever $\hat{\pi}$ takes values above 0.5, no combination of π and q_0 makes a simple scheme sufficient. Numerical robustness checks with respect to the level of effort disutility, e, indicate that a simple scheme is sufficient for a smaller set of $(\pi, \hat{\pi})$ combinations when e is larger; to illustrate this fact, the right panel in Figure 6 plots the sufficiency of a simple scheme for $e = .4$, which is greatly reduced from that of a firm with $e = .1$.

3.3.3 The role of options and refresher grants in implementing the optimal contract

Given that condition b) in Proposition 5 is necessary and sufficient for a complex scheme to be sufficient, but only necessary in the case of a simple scheme, a natural question is whether there is a nontrivial role for complex instruments. Can they help in the implementation of the optimal contract when a simple scheme is not sufficient? The answer to these questions is positive.

The next example illustrates the role of this refresher grant for a specific parametrization of a type $\mathcal{H}$ firm. For the firm in the example, a simple scheme is not sufficient because condition a) is violated. However, condition b) is satisfied, and hence a complex scheme is sufficient: For this firm, the possibility of using a refresher grant s_H allows it to implement the optimal contract without resorting to ad–hoc payments.

Example 2 Consider a firm of type $\mathcal{H}$ described by the following parameters: $\pi = .7$, $\hat{\pi} = .4$, and $q_0 = .3$. The agent has an outside utility of $\bar{U} = 5$, and an effort disutility of $e = .3$.

[22] This property is due to the logarithmic utility and is not likely to be robust for other functional forms of the utility of the agent.

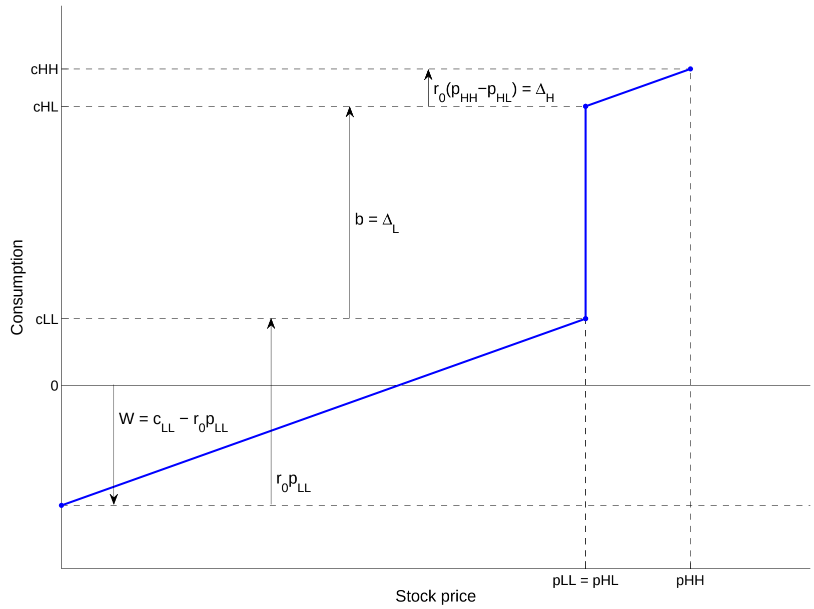

Figure 7: For the firm in Example 2, a simple scheme is not feasible.

(Note that the only difference with the firm in example 1 is in the value of $\hat{\pi}$). The optimal contract is: $c_{LL} = 48.2$, $c_{HL} = 202.3$, and $c_{HH} = 229.3$. The equilibrium stock prices are: $p0 = .79$, $p_L = p_{LL} = p_{LH} = .7$, $p_H = .81$, and $p_{HH} = .84$. For this firm, a simple scheme is not sufficient, as can be seen in Figure 7. Since condition a) fails, the level of r_0 needed to implement Δ_H implies that consumption in the LL and LH states is too high – the only way to hit the target c_{LL} and c_{LH} would be to make W negative – which is not feasible. A complex scheme, instead, is sufficient. One such scheme takes the following values: $W = 48.2$, $b^c = 154.0$, and $s_H = 1039.4$. Figure 8 presents this implementation graphically. A refresher stock grant, s_H, given to the agent after observing a high output in the first period is instrumental in implementing the optimal scheme. Given that there is no variation in prices following a low realization in the first period, the bonus needs to be used to implement Δ_L. Since $\Delta_H < \Delta_L$, the bonus needs to be capped, rather than linear in the accumulated output. Hence, an instrument that will only pay off in the HH history is needed: an option s_H granted when market price is equal to p_H.

The next proposition illustrates that, more generally, there is a nontrivial role for refresher stock grants and option grants.

Proposition 6 *There is a nonempty set of type $\mathcal{H}$ firms for which a simple scheme is not sufficient but a complex one is.*

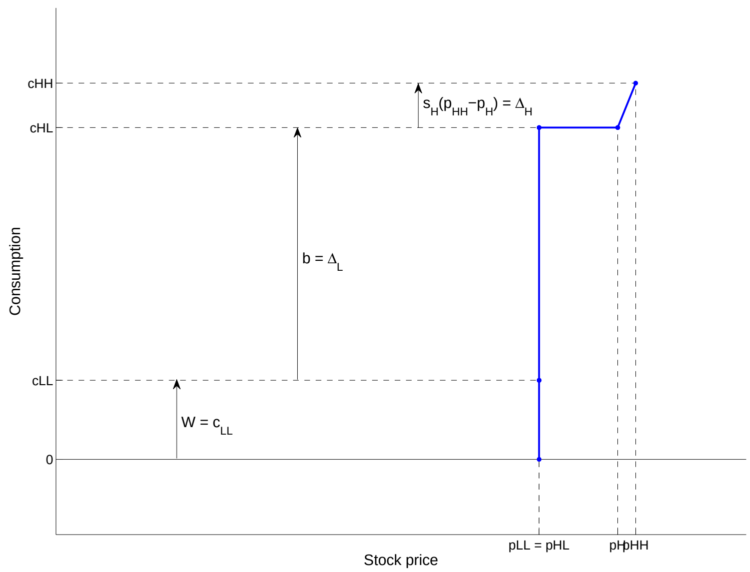

Figure 8: For the firm in Example 2, a complex scheme is feasible.

The proof is by example: Figure 9 presents graphically a grid of the combinations of q_0, π and $\hat{\pi}$ for which a simple scheme is not sufficient but a complex one is. That is, condition a) in Proposition 5 is violated but condition b) is satisfied.

4 Generalization of the model

A more general firm description than the one I have been using as the leading example (the type $\mathcal{H}$ firm) would be one as in matrix 1, where output in state B also depends on the effort choice of the agent. When considering this general case, I assume that $\pi_\theta \neq \widehat{\pi}_\theta$ for at least one θ, and the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H (for any quality of the firm): $\pi_A \geq \widehat{\pi}_A$ and $\pi_B \geq \widehat{\pi}_B$, with at least one being a strict inequality. For a firm of this generality, moral hazard and learning interact in more complicated ways; however, I will argue in this section that the intuition behind the properties of optimal consumption will rely on the same forces highlighted earlier for a type $\mathcal{H}$ firm.

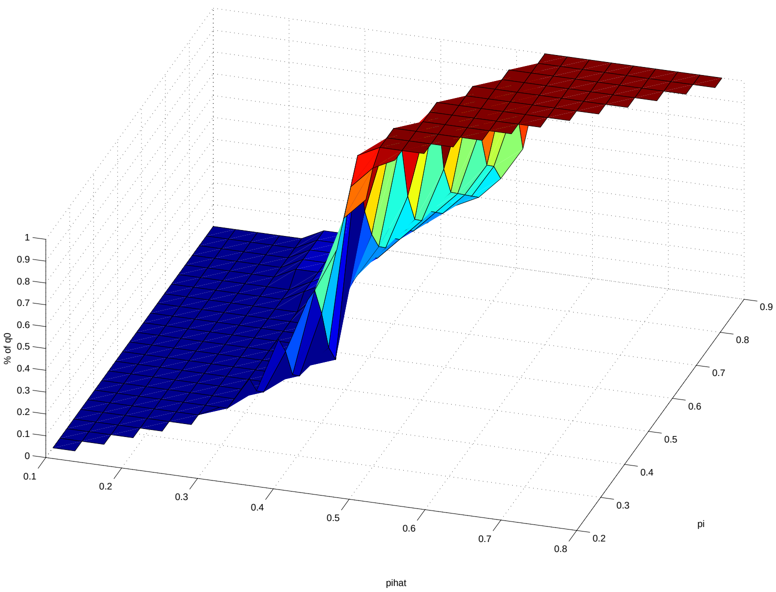

Figure 9: Proportion of firms for which a simple scheme is not feasible but a complex one is (type $\mathcal{H}$ firm).

4.1 A type $\mathcal{L}$ firm

For the sake of this argument, I now introduce a second special type of firm. Let a firm of type $\mathcal{L}$ be described, for all t, by:

$\mathcal{L}$ $\Pr(y_t = y_H \mid e, \theta)$	(q_0) A	$(1-q_0)$ B
e_H	π	0
e_L	$\widehat{\pi}$	0

(10)

In such a firm, when effort is not effective the firm always produces low output, in contrast to the type $\mathcal{H}$ firm, where output is always high in state B. The complete set of counterpart results for type $\mathcal{L}$ firms to the results presented for a type $\mathcal{H}$ firm are included in the appendix. Here, I present the discussion of the results, compare them to those of a type $\mathcal{H}$ firm, and use both sets of results to frame the discussion about the general firm case.

The first thing to note about a firm of type $\mathcal{L}$ is that the properties of consumption in the optimal contract differ from those of a type $\mathcal{H}$ firm: Non–monotonicities never arise following a high realization in the first period ($\Delta_H > 0$ always), but we may have $\Delta_H > \Delta_L$ and non–monotonicities following a low realization in the first period ($\Delta_L < 0$) (see Proposition 11 in the Appendix).

As with a type $\mathcal{H}$ firm, a complete characterization of the sufficiency of simple schemes in terms of primitives is not possible. The next proposition presents the necessary and sufficient conditions, where condition a) is again in terms of the endogenous variable c_{LL}. Figure 10 presents a numerical characterization of the sufficiency of a simple scheme for two different parametrizations of the disutility of effort, e.

Proposition 7 *For a type $\mathcal{L}$ firm, a simple scheme with a capped bonus is never sufficient. A simple scheme with a linear bonus is sufficient if and only if:*

a) $\frac{c^*_{LL}}{\Delta_L - \Delta_H} > \frac{p_{LL}}{p_{LH} - p_{LL}}$, *and*

b) $\Delta_L > \Delta_H$, or $q_0 > q_1^{\mathcal{L}} = \frac{\pi^2(1-\pi)+2\pi\widehat{\pi}-\pi\widehat{\pi}^2-\widehat{\pi}^2}{2\pi\widehat{\pi}(\pi-\widehat{\pi})}$ *(see Proposition 11).*

The insufficiency of a capped bonus is straight forward: Because for an $\mathcal{L}$ firm we have that $p_{HH} = p_{LH}$, it follows that under a capped bonus Δ_H would need to be zero, which is never the case for the nontrivial parametrization that we study, with $\pi \neq \widehat{\pi}$ and $q_0 \in (0,1)$. On the other hand, a linear bonus, defined as $B^l = b^l(y_1 + y_2)$, may be sufficient if the following solution is positive for all three instruments:

$$\begin{aligned} W &= c^*_{LL} - q_{LL}\frac{\Delta_L - \Delta_H}{(1-q_{LL})}, \\ b^l &= \Delta_H, \\ r_0 &= \frac{\Delta_L - \Delta_H}{(1-q_{LL})\,\pi}. \end{aligned}$$

The form of this solution is intuitive. Again, because for an $\mathcal{L}$ firm, $p_{HH} = p_{LH}$, the spread $c_{HH} - c_{LH}$ needs to be implemented through a bonus payout: $b^l = \Delta_H$. Given the linear bonus program, this is not a problem as long as $\Delta_L \geq \Delta_H$. If this is the case, the quantity of restricted stock is determined to satisfy

$$\Delta_L - b^l = r_0\left(p_{LH} - p_{LL}\right).$$

It is the case that $\Delta_L \geq \Delta_H$ whenever incentives need to be more high powered in the low range of outcomes; since an $\mathcal{L}$ firm is more likely to get low output levels out of luck (when the true state is B,) incentives are high powered in the low range of outcomes only for high enough prior that the state is A $\left(\text{i.e., } q_0 > q_1^{\mathcal{L}}\right)$. Finally, it must be the case that the implied wage given r_0 is positive (condition a) in Corollary 7):

$$W = c_{LL} - r_0 p_{LL}.$$

If this condition is not met, a simple scheme is not sufficient. As was the case for type $\mathcal{H}$ firms, the examples in Figure 10 suggest that higher values for effort disutility imply that the sufficiency of a simple scheme is less likely to happen. (Note that by Lemma 3 condition a) in Proposition 7 is independent of the value of $\bar{U}$).

When turning to the implementation with more complex compensation packages, the case $\Delta_H > \Delta_L$ is the one that prompts the use of ad–hoc compensation — so for this firm nontransparent schemes may be needed even if the optimal contract is monotonic in cumulative output. This is summarized in the following characterization of the sufficiency of complex schemes:

Proposition 8 *A complex scheme is sufficient for a type $\mathcal{L}$ firm (with a linear bonus) if and only if $\Delta_L \geq \Delta_H$.*

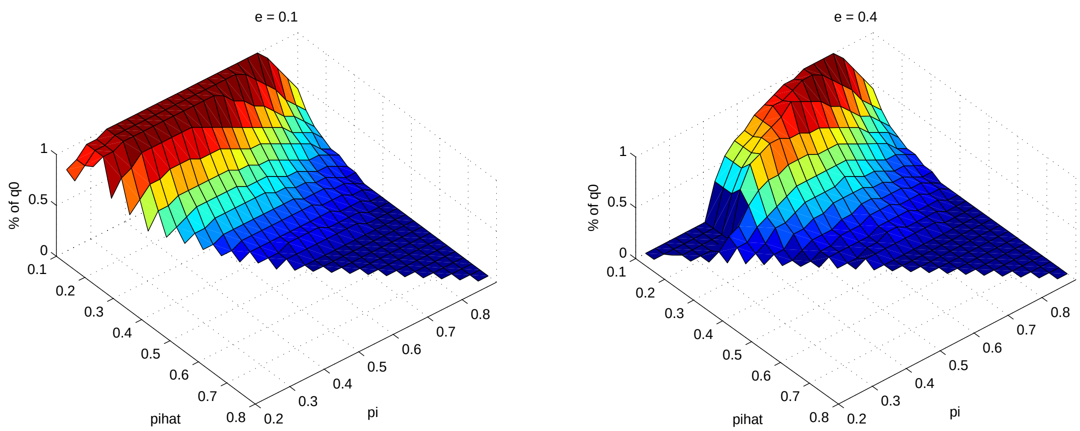

Figure 10: Fraction of type $\mathcal{L}$ firms for which a simple scheme is sufficient, for $e = .1$ (left panel) and $e = .4$ (right panel). There is sufficiency for a smaller set of firms when effort disutility is higher.

As the proof of this proposition shows, complex schemes that are sufficient when $\Delta_L \geq \Delta_H$ also can take a very simple form, such as:

$$\begin{aligned} W &= c_{LL} \\ b^l &= \Delta_H \\ r_H &= \frac{\Delta_L - \Delta_H}{p_{LH}} \\ s_L &= \frac{\Delta_L - \Delta_H}{p_{LH} - p_L} \\ s_0 &= r_0 = r_L = r_0 = 0, \end{aligned}$$

where the bonus is linear in this case. Proposition 11 in Appendix 2 shows that $\Delta_L < \Delta_H$ whenever $q_0 < q_1^{\mathcal{L}}$. We see that for a type $\mathcal{L}$ firm refresher grants may be transferred to the agent even after a bad interim result of the firm (s_L). This was not the case for a type $\mathcal{H}$ firm. The reason is that a type $\mathcal{L}$ firm is successful only when effort is effective, so prices are not sensitive to a second high output realization, while the optimal incentives are; this implies that the bonus should be linear in output in order to implement Δ_H. Because it is still the case that $\Delta_L \geq \Delta_H$, a grant that is only given after a low output and pays only if the second period output is high is instrumental in avoiding ad–hoc payments. The following example illustrates this role of refresher grants for a firm of type $\mathcal{L}$.

Example 3 The following parameters describe an example of a firm of type $\mathcal{L}$ for which a simple scheme is not sufficient, but a complex one is: $\pi = 0.4$, $\hat{\pi} = 0.2$, $q_0 = .8$. The optimal contract is: $c_{LL} = 103.9$, $c_{HL} = 346.5$, and $c_{HH} = 474.1$. The equilibrium stock prices are: $p_0 = .32$, $p_L = .28$, $p_{LL} = .24$, $p_{LH} = p_H = p_{HH} = .4$. For this firm, a complex scheme is sufficient,

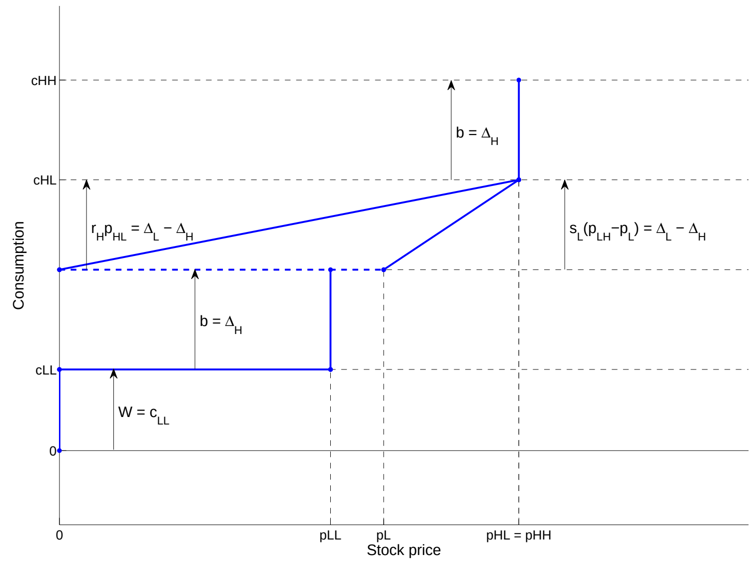

Figure 11: For the type $\mathcal{L}$ firm in example 3, a simple scheme is not feasible but a complex one including refresher grants (r_H and s_L) is feasible.

and it takes the following values: $W = 103.9$, $b^l = 127.6$, $s_L = 977.9$, and $r_H = 287.6$. Figure 11 presents this scheme graphically. The refresher grants are needed to provide $\Delta_L - \Delta_H$. After a first period y_H, this difference is achieved by granting stock r_H, which will pay the same in the state HH as in the state HL (and hence $b^l = \Delta_H$ will implement $c_{HH} - c_{HL}$). After a first period y_L, it is achieved by granting s_L, which has an exercise price of p_L and hence will only pay in the state HL, since $p_{LL} < p_L$ always.

In general, we also find a nontrivial role for complex schemes for type $\mathcal{L}$ firms, as stated in the next proposition.

Proposition 9 *There is a nonempty set of firms (parameter values) for which a simple scheme is not sufficient but a complex one is.*

The proof is by example; for a firm of type $\mathcal{L}$ for $\bar{U} = 5$ and $e = .3$, Figure 12 presents a set of parameters for which a simple scheme is not sufficient but a complex one is.

4.2 Implications for other types of firms

Firms of type $\mathcal{H}$ and $\mathcal{L}$ represent particular examples of technologies that we may identify in real life. For example, we may think of $\mathcal{H}$ firms as mature, successful ones in which only the possibility of a bad match between the CEO and the firm, or an adverse change in technology or in regulations, triggers bad realizations and the need for high effort of the CEO. Type $\mathcal{L}$ firms, instead, may be

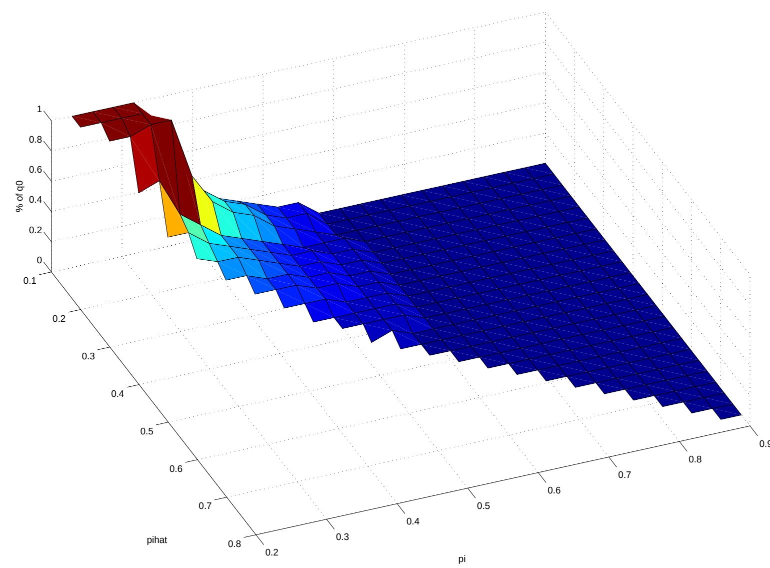

Figure 12: For a type $\mathcal{L}$ firm (for $\bar{U} = 5$ and $e = .3$), sufficiency of a complex scheme whenever a simple one is not sufficient.

younger, struggling firms for which a good match, or favorable change in technology or regulations, need to be paired with high effort to improve outcomes.

Consider now a firm for which, in state B, a high output is realized with probability $\bar{\pi}$, regardless of the effort choice of the agent. Learning interacts with the moral hazard problem in this setting in a more general way than in the type $\mathcal{H}$ and $\mathcal{L}$ firms, since a high output observation is more or less informative about the moral hazard problem depending on the particular value of $\bar{\pi}$. For this firm, if $\bar{\pi}$ is closer to 1 than to 0, stock prices p_L, p_{LH} and p_{LL} will all be close to π_A, but not all equal as they were for a type $\mathcal{H}$ firm. The low sensitivity of prices to output, however, will condition the usage of certain instruments in a similar way to how they did for a type $\mathcal{H}$ firm. For example, using a grant r_0 to implement Δ_L will not be impossible now, but $p_{LL} - p_{LH}$ being small will imply that a very large r_0 will be needed, which may imply too large of a difference in consumption across the LH and HH histories. Again, refresher grants may prove useful.

In the most general type of firm, described in matrix 1, the learning will affect prices in a similar way as we just described for a firm with probability $\bar{\pi}$ across the two effort choices. However, the fact that effort will affect the probability in state B will imply that the optimal contract will have to accommodate for the incentives in state B as well. This will not change the fact that the choice of compensation instruments will still be guided to translate the sensitivity of prices into the sensitivity of consumption, in a similar fashion as the one described for the type $\mathcal{H}$ firm.

It is easy to find examples for which the optimal contract will not be monotonic in output (see Lemma 4 in the Appendix). Because the state HL is the only one that can have a likelihood ratio greater or smaller than 1, the non–monotonicities will only arise following one particular realization of output in the first period (i.e., only one of Δ_L or Δ_H may be negative at a time). In particular, parametrizations for $\pi_A, \hat{\pi}_A, \pi_B$ and $\hat{\pi}_B$ that are close to a type $\mathcal{H}$ firm may have $\Delta_H < 0$, while parametrizations close to a type $\mathcal{L}$ firm may have $\Delta_L < 0$. Following the same logic as the one we developed when studying a type $\mathcal{H}$ firm, we can show that monotonicity will be both necessary and sufficient for a complex scheme to implement the optimal contract.

Proposition 10 *A complex scheme is sufficient for a general type firm (with a capped bonus) if and only if consumption is monotonic in the optimal contract (both $\Delta_L \geq 0$ and $\Delta_H \geq 0$ hold) and $\pi_A \neq \pi_B$.*

Necessity of the monotonicity follows from the fact that any complex scheme implies $\Delta_H \geq 0$ and $\Delta_L \geq 0$. Sufficiency follows from the fact that we can provide a scheme that always implements the optimal contract for a firm of a general type; this is the same scheme in 9, which was used to prove sufficiency in Proposition 4. The knife–edge case $\pi_A = \pi_B$ implies no variation in stock prices, so ad–hoc payments are needed to implement any variation in consumption. This scheme only uses a wage W, a capped bonus, and a refresher grant s_H. Hence, even with the richer stock price dynamics of a general firm, the implementation of the optimal scheme with a complex compensation package needs to use refresher grants.

For a firm of a general type, a numerical characterization of the sufficiency of simple schemes is not easily summarized graphically, since there are two extra parameters compared to the special type firms.[23] Figure 13 presents, in the top and bottom left panel, the sufficiency of simple schemes for firms with the same fixed value of π_A and $\hat{\pi}_A$, for two different pairs of π_A and $\hat{\pi}_A$. In both cases, around the case $\pi_A = \pi_B$ a simple scheme is not feasible. However, there are many other firm specifications for which sufficiency fails. The figures on the right top and bottom panels present the combinations of parameters for which a nontransparent scheme is needed. For the firm in the top panel, we have $\pi_A + \hat{\pi}_A < 1$, and we see that when π_B and $\hat{\pi}_B$ both take values close to 1 these general firms have non–monotonicities in their optimal contract; these are firms similar to a type $\mathcal{H}$ firm, for which $\Delta_L < 0$ when q_0 is low enough (see Proposition 3). For the firms in the bottom panel, we have $\pi_A + \hat{\pi}_A > 1$, and when π_B and $\hat{\pi}_B$ both take values close to 0 these general firms also exhibit non–monotonicities; these are firms similar to a type $\mathcal{L}$ firm, for which $\Delta_H < 0$ when q_0 is low enough (see Proposition 11). As stated in Proposition 10, the right panels show that ad–hoc payments will also be needed for the knife–edge case of $\pi_A = \pi_B$.

Again, we find that complex instruments fulfill a nontrivial role, since for some firms a simple scheme is not sufficient, but a complex one is. The next example discusses such a firm.

Example 4 The following parameters describe an example of a firm of a general type for which a simple scheme is not sufficient, but a complex one is: $\pi_A = 0.9$, $\hat{\pi}_A = 0.7$, $\pi_B = 0.3$, $\hat{\pi}_B = 0.1$, $q_0 = .8$, $e = .075$, $\bar{U} = 5$. The optimal contract is: $c_{LL} = 122.8$, $c_{HL} = 141.1$, and $c_{HH} = 174.2$. The equilibrium stock prices are: $p_0 = .78$, $p_L = .52$, $p_{LL} = .35$, $p_{LH} = .68$,

[23] See Propositions 13 and 12 in the Appendix for the three conditions that need to be satisfied for a simple scheme to be sufficient for a general type firm, for a capped and a linear bonus scheme, respectively.

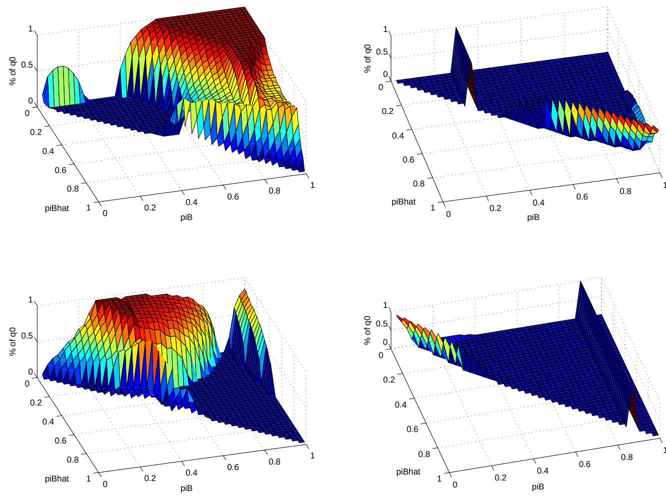

Figure 13: The top two panels correspond to firms with $\pi_A = .3$ and $\hat{\pi}_A = .1$. The bottom two panels correspond to firms with $\pi_A = .9$ and $\hat{\pi}_A = .5$. For all firms, $\bar{U} = 5$ and $e = .2$. The left panels present the percentage of firms for which a simple scheme is sufficient. The right panels present the percentage of firms for which nontransparent schemes are necessary.

$p_H = .85$, $p_{HH} = .88$. For this firm, a complex scheme is sufficient, and it takes the following values: $W = 122.8$, $b^c = 18.3$, $s_H = 1103.7$. Figure 14 presents this scheme graphically. The refresher stock grant is needed to provide Δ_H, since trying to use r_0 to stick to a simple scheme would imply too high consumption in the state HL (and hence a negative capped bonus would be necessary, which violates the non–negativity constraint).

4.3 Empirical evidence on refresher grants

With the implications for a general type firm in mind, it is worth mentioning some interesting — although scarce — evidence about one particular complex compensation practice: refresher grants.[24] Hall and Knox (2004) find evidence that larger than average grants are often used by firms both following a stock price decline and following a stock price increase.[25] They interpret refresher grants as a mechanism to restore incentives for the CEO whenever the sensitivity of his compensation to

[24] The related practice of option "repricing" is fairly uncommon (Brenner, Sundaram, and Yermak, 2000), and not important in adjusting incentives over time (Hall and Knox, 2004). For studies of the explanatory power of observable characteristics on the probability of repricing see also Chance, Kumar, and Todd (2000), Brenner, Sundaram and Yermak (2000), Carter and Lynch (2001), and Chen (2004).

[25] Hall (1999) uses a more inclusive definition of refresher grants (out–of–cycle or larger than average), while Hall and Knox (2004) call refresher grants only out–of–cycle grants (in particular, within the same fiscal year in which a large change in the stock price is observed). In their evidence, they find the importance of larger–than–usual grants to be much greater than that of out–of–cycle grants. For this discussion I use the more comprehensive definition.

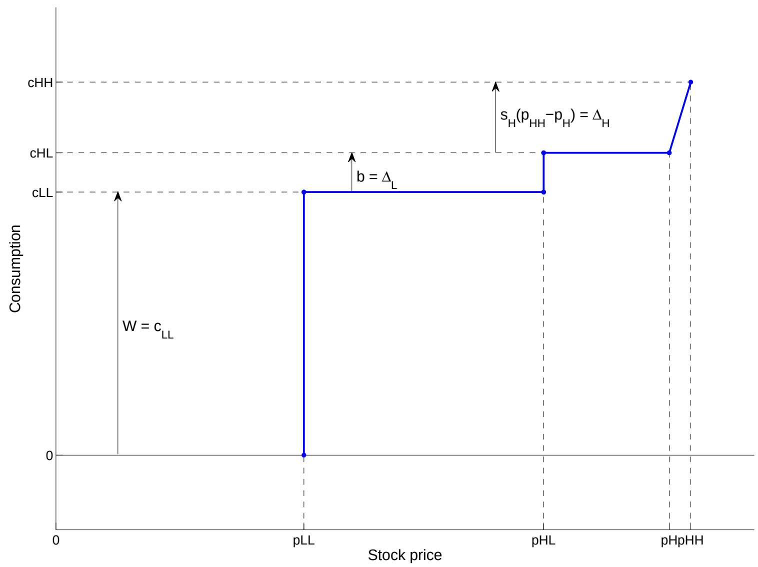

Figure 14: For the general type firm in Example 4, a complex scheme with a capped bonus is feasible.

stock price movements decreases, and report that refresher grants following a stock price increase seem puzzling under this logic (since stock price increases tend to increase the sensitivity of pay to firm performance).[26] My model provides a rationale for new grants contingent on both good and bad firm performance.

5 Conclusion

In this paper, I ask the question of what are the firm characteristics that may justify the use of options, refresher grants, or ad–hoc payments such as signing bonuses or discounted purchases of stock in the compensation packages for CEOs. I view compensation packages as particular implementations of the optimal contract that provides incentives to the CEO to exert high effort at a minimum cost in the presence of moral hazard. Working with models of asymmetric information and risk–averse agents is generally difficult. Here, I present a necessarily stark model of a firm. Its simplicity allows me to enrich it with learning about the effectiveness of the effort of the CEO in enhancing the output of the firm with his or her effort. This provides me with a model that explains stock prices and compensation jointly from primitives. One lesson emerges from the analysis of the model: The level of uncertainty about (and the priors on) the effectiveness of the CEO's actions is

[26] Core and Guay (1999), without trying explicitly to establish whether a given grant is a refresher grant (out of the ordinary in some dimension), find evidence that firms increase or decrease their annual incentive grants to maintain an optimal incentive level. The results in this paper are consistent with their findings as well.

an important factor for the type of compensation instruments that the firm uses.

The results show that a limited set of compensation instruments such as an uncontingent wage, a bonus plan linear on output, and stock grants, may be too simple to implement the optimal contract. Still, for a nontrivial set of firms this is sufficient. On the other hand, more complex schemes including also refresher stock or option grants (contingent on new information about the performance of the firm) are generally sufficient. Such a complex scheme is not sufficient when the learning about the effectiveness of the effort of the CEO is very important, which will typically imply that the optimal compensation contract is non–monotonic in output. For this class of firms, ad–hoc payments are necessary to provide incentives in the most cost–efficient way.

6 Appendix

Proof of Lemma 2. The weak monotonicity of prices follows from

$$p\left(y^t, e\right) \equiv E_t\left[y_t \mid e\right],$$

the fact that, for each firm, either $\pi_A > \pi_B$ or $\pi_B > \pi_A$ is true. Then, by Bayes' rule, the posterior of the state with higher probability is weakly increasing in the number of high output realizations. ■

Proof of Proposition 1. With $u(c) = \ln(c)$, the non–negativity constraints on consumption will not bind. With $\lambda \geq 0$ as the multiplier for the binding PC, and $\mu \geq 0$ for the binding IC, the first order conditions of the problem are

$$\frac{1}{u'(c_{ij})} = \lambda + \mu(1 - LR_{ij}),$$

which imply the ranking because $u'(\cdot)$ is monotonically decreasing in consumption. These simplify to equation 8 in the case of $u(c) = \ln(c)$. ■

Proof of Proposition 2. It is easy to see this by looking at the likelihood ratios in this particular case of our framework, which simplify to:

$$\begin{aligned} LR_{LL} &= \frac{1-\widehat{\pi}}{1-\pi}\frac{1-\widehat{\pi}}{1-\pi} \\ LR_{LH} &= \frac{1-\widehat{\pi}}{1-\pi}\frac{\widehat{\pi}}{\pi} \\ LR_{HL} &= \frac{\widehat{\pi}}{\pi}\frac{1-\widehat{\pi}}{1-\pi} \\ LR_{HH} &= \frac{\widehat{\pi}}{\pi}\frac{\widehat{\pi}}{\pi}, \end{aligned}$$

where $\frac{\widehat{\pi}}{\pi} < 1 < \frac{1-\widehat{\pi}}{1-\pi}$ and hence $LR_{HH} < LR_{HL} = LR_{LH} < LR_{LL}$, which implies $c_{HH} > c_{HL} = c_{LH} > c_{LL}$. Moreover,

$$\begin{aligned} \Delta_H &= \mu\left(LR_{LH} - LR_{HH}\right) = \mu\frac{\widehat{\pi}}{\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)}, \\ \Delta_L &= \mu\left(LR_{LL} - LR_{LH}\right) = \mu\frac{1-\widehat{\pi}}{1-\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)}. \end{aligned}$$

Since μ is the Lagrange multiplier of the incentive constraint in problem (**P1**), it will satisfy $\mu \geq 0$. Since $\pi > \widehat{\pi}$ by assumption, the second result in the proposition follows. ■

Proof of Proposition 3. In the first part of the proposition we want to show that, for a firm of type $\mathcal{H}$, $\Delta_H < \Delta_L$, and also that $\Delta_L > 0$. We first show the second inequality holds, and then we use it to prove the first inequality. From the expressions for the likelihood ratios in 7, we have

$$\begin{aligned}\Delta_L &= \mu\left(LR_{LL} - LR_{LH}\right) \\ &= \mu\frac{1-\widehat{\pi}}{1-\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)}.\end{aligned}$$

Since μ is the Lagrange multiplier of the incentive constraint in problem (**P1**), it will satisfy $\mu \geq 0$. Since $\pi > \widehat{\pi}$ by assumption, $\Delta_L > 0$ follows. Now to establish $\Delta_H < \Delta_L$ we first note that

$$\frac{\Delta_L - \Delta_H}{\mu} = \left(LR_{LL} - LR_{LH}\right) - \left(LR_{LH} - LR_{HH}\right),$$

where all likelihood ratios except for LR_{HH} are independent of q_0. As we showed in the proof of Proposition 2, we have that $LR_{HL} = LR_{LH} < LR_{LL}$. In turn, we can show that LR_{HH} decreases monotonically for $q_0 \in (0,1)$:

$$\frac{\partial LR_{HH}}{\partial q_0} = \frac{\widehat{\pi}^2 - \pi^2}{\left(q_0\pi^2 + 1 - q_0\right)^2} < 0.$$

Then, by taking the limit of LR_{HH} with respect to q_0 we can bound $\frac{\Delta_H}{\mu}$ and compare it to $\frac{\Delta_L}{\mu}$ in both extreme cases. When q_0 approaches 0, LR_{HH} goes to its maximum, 1, and we have:

$$\lim_{q_0 \to 0} \frac{\Delta_H}{\mu} = \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi} - 1,$$

which determines the minimum possible $\frac{\Delta_H}{\mu}$. When q_0 approaches 1, instead, LR_{HH} approaches its minimum, which coincides with the expression for LR_{HH} in the no learning case, $\frac{\widehat{\pi}^2}{\pi^2}$:

$$\lim_{q_0 \to 1} \frac{\Delta_H}{\mu} = \frac{\widehat{\pi}}{\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)}.$$

This is the maximum possible value of $\frac{\Delta_H}{\mu}$. For this maximum value of $\frac{\Delta_H}{\mu}$, we have

$$\frac{\Delta_L - \Delta_H}{\mu} = \frac{1-\widehat{\pi}}{1-\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)} - \frac{\widehat{\pi}}{\pi}\frac{\pi-\widehat{\pi}}{\pi(1-\pi)},$$

which again is positive since $\pi > \widehat{\pi}$ by assumption. This implies $\Delta_H < \Delta_L$ for any value of q_0. Note that μ depends on q_0, but the proof works because the comparison is for a given common μ in both Δ_H and Δ_L. For the second part of the proposition, note that we have that $\Delta_H < 0$ if and only if

$$\frac{\hat{\pi}\,(1-\hat{\pi})}{\pi\,(1-\pi)} < \frac{q_0\hat{\pi}^2 + 1 - q_0}{q_0\pi^2 + 1 - q_0}.$$

Rearranging, this condition becomes, for $\hat{\pi} + \pi < 1$,

$$q_0 < \frac{1 - \pi - \hat{\pi}}{1 - \pi - \hat{\pi} + \pi\hat{\pi}}.$$

Whenever $\pi + \widehat{\pi} \geq 1$, the inequality is never satisfied for a $q_0 \in (0, 1)$. ■

Proof of Proposition 4. (Necessity) It is useful to write the function $\mathcal{C}(P)$ in 5 as a function of the consumption differences Δ_L and Δ_H defined previously, as well as using the following notation for the price differences: $\alpha_L = p_{LH} - p_{LL}$ and $\alpha_H = p_{HH} - p_{LH}$. For a capped bonus and $p_{LH} < p_0$, for example, the system becomes:

$$\begin{aligned}
c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\
\Delta_H &= r_0 \alpha_H + s_0 (p_{HH} - p_0) + r_H \alpha_H + s_H (p_{HH} - p_H) \\
\Delta_L &= b + r_0 \alpha_L + r_L \alpha_L + s_L (p_{HL} - p_L) \\
0 &= r_H p_{HL} - r_L p_{HL} - s_L (p_{HL} - p_L).
\end{aligned}$$

The fourth restriction comes from the property of the optimal contract that states that $c_{LH} = c_{HL}$. It is easy to see from this system that no values for the instruments will ever be able to implement $\Delta_H < 0$, since all $\alpha' s$ are positive and $p_{HH} > p_H$ always. The argument is similar for the case of $p_{LH} \geq p_0$, and the corresponding two cases with a linear bonus. (Sufficiency). The following scheme with a capped bonus is always a solution to the system $C^* = \mathcal{C}(P)$:

$$\begin{aligned}
W &= c_{LL} \\
b &= \Delta_L \\
s_H &= \frac{\Delta_H}{p_{HH} - p_H} \\
s_0 &= r_0 = r_L = r_H = s_L = 0.
\end{aligned}$$

Note that because the solution has $s_0 = 0$ it implements the same consumption regardless of whether p_{LH} is greater or smaller than p_0. ■

Proof of Proposition 5. For a type $\mathcal{H}$ firm, it is straightforward to see that a linear bonus is insufficient, since we have $p_{LL} = p_{LH}$, or $\alpha_L = 0$. This means that the spread Δ_L must be implemented with the bonus b. However, a linear bonus implies $\Delta_H \geq b$, while we saw in Proposition 3 that for this type of firm it is always the case that $\Delta_H < \Delta_L$. Given this result, the proposition is a corollary of Proposition 13. Note that the condition $\frac{\Delta_L}{\Delta_H} \geq \frac{p_{LH} - p_{LL}}{p_{HH} - p_{LH}}$ is implied by the condition $\Delta_H \geq 0$ for an $\mathcal{H}$ firm, since $p_{LH} - p_{LL} = 0$ for this type of firm. ■

Lemma 3 *When $u(c) = \ln(c)$, consumptions in the optimal contract change proportionally with changes in $\bar{U}$.*

Proof of Lemma 3. Denote by c^*_{ij}, for $i, j = L, H$, the solutions to problem PS when the outside utility of the agent is equal to $\bar{U}$. We want to show that $c_{ij} = \alpha c^*_{ij}$, where $\alpha = \exp(a)$, is a solution to problem PS when all other parameters remain the same but the outside utility of the agent is equal to $\bar{U}' = \bar{U} + a$. We will do this by showing that this proposed solution satisfies the PC, the IC and the first order conditions for the problem with parameter $\bar{U}'$. With a slight abuse of notation, let π_{ij} and $\hat{\pi}_{ij}$ denote the probabilities of history (y_i, y_j) on and off the equilibrium path. First, the PC is

$$\bar{U}' = \sum_{ij} \pi_{ij} \ln(c_{ij}) - e.$$

Using the proposed solution, this becomes:

$$\begin{aligned}
\bar{U}' &= \sum_{ij} \pi_{ij} \ln\left(\alpha c_{ij}^*\right) - e, \\
\bar{U} + a &= \ln \alpha + \sum_{ij} \pi_{ij} \ln\left(c_{ij}^*\right) - e,
\end{aligned}$$

which is true since $a = \ln \alpha$ and c_{ij}^* satisfies the PC when the outside utility is $\bar{U}$. Second, the IC is

$$\sum_{ij} (\pi_{ij} - \hat{\pi}_{ij}) \ln(c_{ij}) = e.$$

With the proposed solution this becomes

$$\begin{aligned}
\sum_{ij} (\pi_{ij} - \hat{\pi}_{ij}) \ln\left(\alpha c_{ij}^*\right) &= e, \\
\sum_{ij} (\pi_{ij} - \hat{\pi}_{ij}) \ln \alpha + \sum_{ij} (\pi_{ij} - \hat{\pi}_{ij}) \ln\left(c_{ij}^*\right) &= e, \\
\sum_{ij} (\pi_{ij} - \hat{\pi}_{ij}) \ln\left(c_{ij}^*\right) &= e,
\end{aligned}$$

which is the IC for the problem with $\bar{U}$, and hence it is satisfied by c_{ij}^*. Finally, the first order conditions are

$$c_{ij} = \lambda' + \mu' (1 - LR_{ij}), \text{ for } i, j = L, H,$$

where λ and μ are the lagrange multipliers of the problem with $\bar{U}'$.Using the proposed solution we can see that $\lambda = \alpha\lambda$ and $\mu' = \alpha\mu$ (where λ and μ are the lagrange multipliers of the problem with $\bar{U}$) satisfy all the first order conditions as well when $c_{ij} = \alpha c_{ij}^*$. Hence, the proposed solution is a true solution. ■

6.1 Generalizations

6.1.1 Type $\mathcal{L}$ firm prices and consumption properties

Consider instead a type $\mathcal{L}$ firm, described by matrix (10). The analysis of this case parallels that of case $\mathcal{H}$. For any histories containing at least one y_H, the updated believes put probability one on $\theta = A$, i.e., we have that $q_{ij}^{\mathcal{L}} = 1$ if i or j equals H. If the observed history does not contain any y_H, instead, $\theta = B$ has still positive probability. This is the case for histories y_L and (y_L, y_L). That is,

$$\begin{aligned} q_L^{\mathcal{L}} &= q_0 \frac{(1-\pi)}{q_0(1-\pi)+1-q_0}, \\ q_{LL}^{\mathcal{L}} &= q_0 \frac{(1-\pi)^2}{q_0(1-\pi)^2+1-q_0}, \\ q_H^{\mathcal{L}} &= q_{LH}^{\mathcal{L}} = q_{HH}^{\mathcal{L}} = 1. \end{aligned}$$

The stock prices take the simple form:

$$\begin{aligned} p_0^{\mathcal{L}} &= q_0 \pi, \qquad (11) \\ p_L^{\mathcal{L}} &= q_L^{\mathcal{L}} \pi, \\ p_{LL}^{\mathcal{L}} &= q_{LL}^{\mathcal{L}} \pi, \\ p_H^{\mathcal{L}} &= p_{HL}^{\mathcal{L}} = p_{HH}^{L} = \pi. \end{aligned}$$

For this type of firm we have that $q_H^{\mathcal{L}} = q_{LH}^{\mathcal{L}} = q_{HH}^{\mathcal{L}} = 1$. Learning may in this case also give rise to non–monotonicities. In particular, when the first period output has been y_L, the agent's wage may be lower if we observe y_H in the second period than if we observe y_L. This is because observing y_H in the second period reveals that $\theta = A$. In state A, the first period observation y_L makes the history's likelihood ratio be much higher (it is a much more likely history under low effort than it would be if $\theta = B$). Formally, the likelihood ratios are:

$$\begin{aligned} LR_{LL}^{\mathcal{L}} &= \frac{q_0(1-\widehat{\pi})^2+1-q_0}{q_0(1-\pi)^2+1-q_0} \qquad (12) \\ LR_{LH}^{\mathcal{L}} &= \frac{(1-\widehat{\pi})\widehat{\pi}}{(1-\pi)\pi} \\ LR_{HH}^{\mathcal{L}} &= \frac{\widehat{\pi}^2}{\pi^2}. \qquad (13) \end{aligned}$$

The likelihood ratios $LR_{HH}^{\mathcal{L}}$ and $LR_{LH}^{\mathcal{L}}$ coincide in this case with the corresponding ones in the benchmark case of no learning characterized in Proposition 2, when the true state is known to be A. We can use these likelihood ratios to establish the following properties of consumption in the optimal contract:

Proposition 11 *When the firm is of type $\mathcal{L}$ as described by matrix (10), consumption spread satisfies:*

(i) *whenever $\pi + \widehat{\pi} > 1$,*

$$\begin{aligned} \textit{for } q_0 &\in (0, q_1^{\mathcal{L}}],\ \Delta_L < 0 < \Delta_H, \\ \textit{for } q_0 &\in \left[q_1^{\mathcal{L}}, q_2^{\mathcal{L}}\right], 0 < \Delta_L < \Delta_H, \\ \textit{for } q_0 &\in [q_2^{\mathcal{L}}, 1),\ 0 < \Delta_H < \Delta_L, \end{aligned}$$

(ii) *whenever* $\pi + \widehat{\pi} \leq 1$,

$$\begin{aligned} \textit{for } q_0 &\in (0, \max\left\{q_2^{\mathcal{L}}, 0\right\}],\ 0 < \Delta_L < \Delta_H, \\ \textit{for } q_0 &\in [\max\left\{q_2^{\mathcal{L}}, 0\right\}, 1),\ 0 < \Delta_H < \Delta_L. \end{aligned}$$

where

$$\begin{aligned} q_1^{\mathcal{L}} &= \frac{\pi + \widehat{\pi} - 1}{\pi\widehat{\pi}}, \\ q_2^{\mathcal{L}} &= \frac{2\pi\widehat{\pi} - \pi\widehat{\pi}^2 - \widehat{\pi}^2 - \pi^2(1-\pi)}{2\pi\widehat{\pi}(\pi - \widehat{\pi})}. \end{aligned}$$

(The proof is included after this discussion.) Again, we find a difference depending on whether $LR_{LH}^{\mathcal{L}}$ is greater or smaller than one. When $\pi + \widehat{\pi} > 1$, we have that $LR_{LH}^{\mathcal{L}}$ is greater than one and hence the contract seeks to punish the agent at LH as well as at LL. When the firm is of type $\mathcal{L}$, however, observing a high realization implies that the state is A with probability one. This makes a low realization a very valuable (and negative) signal about performance, making c_{LH} but not c_{LL} low. This implies a small Δ_L. For low enough prior of being in state A, this can lead to $\Delta_L < 0$. Higher priors diminish the relative informativeness of LH with respect to LL and hence reestablish the relationship $0 < \Delta_H < \Delta_L$ of the no learning benchmark.

When $\pi + \widehat{\pi} < 1$, we have that $LR_{LH}^{\mathcal{L}}$ is smaller than one and hence the contract seeks to reward the agent at LH as well as at HH. The posterior becomes one under either realization, so the standard ranking of $c_{LH} < c_{HH}$ prevails, implying $\Delta_H > 0$ always. However, since the only state in which there is punishment is LL and this is a very likely outcome when the true state is B, unrelated to effort choice, a small q_0 (a high prior that we are in B) implies a very high cost of utility imposed on the agent with very little incentive benefit (he consumes very low very often but his incentives are little changed, since LL is very unlikely in state A). This tends to keep c_{LL} not too far from c_{LH}, and hence it can lead to $\Delta_L < \Delta_H$. However, for $\pi + \widehat{\pi}$ smaller than one but with $(\pi - \widehat{\pi})$ small the informativeness of signals in state A decreases a lot; this makes c_{HH} close to c_{LH} as well, and more so than c_{LH} close to c_{LL} for the same reasons as in the benchmark case without learning. This is reflected in the case $q_2^{\mathcal{L}} < 0$, which is more likely when $(\pi - \widehat{\pi})$ is small and implies $0 < \Delta_H < \Delta_L$ always.

Proof of Proposition 11. From the expressions for the likelihood ratios in equation 12, we have

$$\begin{aligned} \Delta_H &= \mu\left(LR_{LH}^{\mathcal{L}} - LR_{HH}^{\mathcal{L}}\right) \\ &= \mu\frac{\widehat{\pi}}{\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)}. \end{aligned}$$

The expression for Δ_L cannot be easily simplified. However, by taking the limit of $\frac{\Delta_L}{\mu}$ with respect to q_0, we can bound it and compare it to $\frac{\Delta_H}{\mu}$ in both extreme cases. The only likelihood ratio in the difference that depends on q_0 is LR_{LL}, and it does so monotonically for $q_0 \in (0,1)$:

$$\frac{\partial LR_{LL}^{\mathcal{L}}}{\partial q_0} = \frac{(1-\widehat{\pi})^2 - (1-\pi)^2}{\left[q_0(1-\pi)^2 + 1 - q_0\right]^2} > 0.$$

When q_0 approaches 1, $LR^{\mathcal{L}}_{LL}$ approaches its maximum, which coincides with the expression for LR_{LL} in the no learning case, $\frac{(1-\widehat{\pi})^2}{(1-\pi)^2}$. Note that μ depends on q_0 but it is always positive and its value affects both Δ_L and Δ_H proportionally, so it does not affect the ranking of these two spreads of consumption. Then it is easy to see that:

$$\lim_{q_0 \to 1} \frac{\Delta_L}{\mu} = \frac{1-\widehat{\pi}}{1-\pi} \frac{\pi - \widehat{\pi}}{\pi(1-\pi)} > \frac{\Delta_H}{\mu} > 0.$$

When q_0 approaches 0 instead, $LR^{\mathcal{L}}_{LL}$ goes to its minimum, 1, and we have:

$$\lim_{q_0 \to 0} \frac{\Delta_L}{\mu} = 1 - \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi}.$$

This is not conclusive; we need to analyze two cases separately, according to whether π and $\widehat{\pi}$ are such that *(i) or (ii)* is satisfied. First, we have that $\Delta_L < 0$ if and only if

$$\frac{q_0\,(1-\hat{\pi})^2 + (1-q_0)}{q_0\,(1-\pi)^2 + (1-q_0)} < \frac{\hat{\pi}\,(1-\hat{\pi})}{\pi\,(1-\pi)},$$

or

$$q_0 < q^{\mathcal{L}}_1 \equiv \frac{\hat{\pi} + \pi - 1}{\pi\hat{\pi}}.$$

When $\pi + \widehat{\pi} \leq 1$, there is no value of q_0 for which $\Delta_L < 0$. Note also that $q^{\mathcal{L}}_1 \leq 1$ for any values of π and $\widehat{\pi}$, since:

$$\begin{aligned} \frac{\hat{\pi} + \pi - 1}{\pi\hat{\pi}} &\leq 1 \\ \hat{\pi} + \pi - 1 &\leq \pi\hat{\pi} \\ \pi\,(1-\hat{\pi}) &\leq 1 - \hat{\pi} \\ \pi &\leq 1, \end{aligned}$$

which is always true. For the second threshold, we have that $\Delta_L \leq \Delta_H$ if and only if

$$\frac{q_0\,(1-\hat{\pi})^2 + (1-q_0)}{q_0\,(1-\pi)^2 + (1-q_0)} - \frac{\hat{\pi}\,(1-\hat{\pi})}{\pi\,(1-\pi)} \leq \frac{\widehat{\pi}}{\pi} \frac{\pi - \widehat{\pi}}{\pi(1-\pi)}.$$

Simplifying, we get:

$$q_0 \leq q^{\mathcal{L}}_2 \equiv \frac{2\pi\widehat{\pi} - \widehat{\pi}^2\,(\pi+1) - \pi^2\,(1-\pi)}{2\pi\widehat{\pi}\,(\pi - \widehat{\pi})}.$$

Note that $q^{\mathcal{L}}_2 < 1$ for any combination of probabilities, since the denominator is larger than the numerator:

$$2\pi\widehat{\pi}\,(\pi - \widehat{\pi}) > 2\pi\widehat{\pi} - \widehat{\pi}^2\,(\pi+1) - \pi^2\,(1-\pi)\,,$$

or

$$\begin{aligned} 2\pi^2\widehat{\pi} - \pi\widehat{\pi}^2 - 2\pi\widehat{\pi} + \widehat{\pi}^2 + \pi^2 - \pi^3 &> 0 \\ 2\pi\widehat{\pi}(\pi - 1) - \widehat{\pi}^2(\pi-1) - \pi^2(\pi-1) &> 0 \\ (1-\pi)(\pi^2 - 2\pi\widehat{\pi} + \widehat{\pi}^2) &> 0 \\ (1-\pi)(\pi - \widehat{\pi})^2 &> 0. \end{aligned}$$

However, whether $q_2^{\mathcal{L}}$ is strictly positive depends on the sum of π and $\widehat{\pi}$:

$$\begin{aligned} q_2^{\mathcal{L}} &> 0 \\ 2\pi\widehat{\pi} - \widehat{\pi}^2(\pi+1) - \pi^2(1-\pi) &> 0 \\ 2\pi\widehat{\pi} - \widehat{\pi}^2 - \pi^2 &> \widehat{\pi}^2\pi - \pi^3 \\ \left(\pi^2 - \widehat{\pi}^2\right)\pi &> (\pi - \widehat{\pi})^2 \\ \pi(\pi - \widehat{\pi})(\pi + \widehat{\pi}) &> (\pi - \widehat{\pi})^2 \\ \pi(\pi + \widehat{\pi}) &> \pi - \widehat{\pi}. \end{aligned}$$

If $\pi + \widehat{\pi} \geq 1$, then the last inequality is always satisfied, and hence $0 < q_2^{\mathcal{L}} < 1$. If $\pi + \widehat{\pi} < 1$, however, for some pairs of π and $\widehat{\pi}$ there will be no q_0 for which $\Delta_L \leq \Delta_H$. ■

Proof of Proposition 7. For a type $\mathcal{L}$ firm, it is straightforward to see that a capped bonus is insufficient, since we have $p_{HH} = p_{LH}$, or $\alpha_H = 0$. This means that the spread Δ_H (which is always positive, by Proposition 11) must be implemented with the bonus b, and hence we cannot have it capped. Given this result, the proposition is a corollary of Proposition 12 below, where I consider a linear bonus program: $B^l(y_1, y_2) = b^l(y_1 + y_2)$. Note that the conditions $\frac{\Delta_H}{\Delta_L - \Delta_H} > \frac{p_{HH} - p_{LH}}{(p_{LH} - p_{LL}) - (p_{HH} - p_{LH})}$ and $\frac{\Delta_L - \Delta_H}{(p_{LH} - p_{LL}) - (p_{HH} - p_{LH})} > 0$ for a general firm both simplify to the condition $\Delta_L - \Delta_H > 0$ for an $\mathcal{L}$ firm, since $p_{HH} - p_{HL} = 0$. ■

Proof of Proposition 8. (Necessity) The argument parallels that of the proof of Proposition 4. For a firm of type $\mathcal{L}$, we have

$$\begin{aligned} c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\ \Delta_H &= b^l \\ \Delta_L &= b^l + r_0\alpha_L + s_0(p_{HL} - p_0) + r_L\alpha_L + s_L(p_{HL} - p_L) \\ 0 &= r_H p_{HL} - r_L p_{HL} - s_L(p_{HL} - p_L). \end{aligned}$$

and it is clear that $\Delta_H \leq \Delta_L$ for any combination of instruments with $b^l > 0$, which is needed to implement $\Delta_H > 0$. (Sufficiency) The following solution implements any optimal consumption for a type $\mathcal{L}$ firm:

$$\begin{aligned} W &= c_{LL} \\ b^l &= \Delta_H \\ r_H &= \frac{\Delta_L - \Delta_H}{p_{LH}} \\ s_L &= \frac{\Delta_L - \Delta_H}{p_{LH} - p_L} \\ s_0 &= r_0 = r_L = r_0 = 0. \end{aligned}$$

■

6.1.2 Generalization of results for sufficiency of simple and complex schemes

In this appendix I present the sufficiency results for a general firm as defined, at all t, by matrix 1, where I assume that $\pi_\theta \neq \widehat{\pi}_\theta$ for at least one θ, and the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H (for any quality of the firm): $\pi_A \geq \widehat{\pi}_A$ and $\pi_B \geq \widehat{\pi}_B$, with at least one being a strict inequality.

Lemma 4 *Optimal consumption is not necessarily monotonic in output, i.e. we may have $\Delta_L < 0$ or $\Delta_H < 0$. Also, both $\Delta_H < \Delta_L$ and $\Delta_H > \Delta_L$ may occur.*

The proof for this result, included in Miller (1999), simply analyzes the possible ranking of likelihood ratios for different combinations of probabilities in matrix 1.[27]

First, I study the case of a simple scheme with a linear bonus.

Proposition 12 *A simple scheme with a **linear bonus** is sufficient if and only if:*

a) $\frac{c^*_{LL}}{\Delta_L - \Delta_H} \geq \frac{p_{LL}}{(p_{LH}-p_{LL})-(p_{HH}-p_{LH})}$,

b) $\frac{\Delta_L - \Delta_H}{(p_{LH}-p_{LL})-(p_{HH}-p_{LH})} \geq 0$,

c) $\frac{\Delta_H}{\Delta_L - \Delta_H} \geq \frac{p_{HH}-p_{LH}}{(p_{LH}-p_{LL})-(p_{HH}-p_{LH})}$.

Proof of Proposition 12. With a linear bonus, the function $\mathcal{C}(P)$ implies:

$$\begin{aligned} c^*_{HH} &= W + 2b + r_0 p_{HH} \\ c^*_{HL} &= W + b + r_0 p_{HL} \\ c^*_{LH} &= W + b + r_0 p_{HL} \\ c^*_{LL} &= W + r_0 p_{LL}. \end{aligned}$$

It is useful to write these equations as a function of the consumption differences Δ_L and Δ_H defined in the previous section. To simplify, I also introduce the following notation for the differences in prices: $\alpha_L \equiv p_{LH} - p_{LL}$ and $\alpha_H \equiv p_{HH} - p_{HL}$.

$$\begin{bmatrix} \Delta_H \\ \Delta_L \\ c^*_{LL} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \alpha_H \\ 0 & 1 & \alpha_L \\ 1 & 0 & p_{LL} \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \end{bmatrix}.$$

The solution to this system is:

$$\begin{aligned} W &= c^*_{LL} - p_{LL}\frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}, \\ b &= \Delta_H - \alpha_H\frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}, \\ r_0 &= \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}. \end{aligned}$$

[27] See also Celentani and Loveira (2006) for a parallel result in a model of simultaneous (rather than sequential) signals that can explain the apparent lack of relative performance in executive compensation.

The conditions a)-c) are then derived from the non–negativity constraints imposed on W, b and r_0. ■

A similar analysis can be pursued for the case of a capped bonus, for a general firm as described in matrix 1.

Proposition 13 *A simple scheme with a* ***capped bonus*** *is sufficient if and only if:*

a) $\frac{c_{LL}}{\Delta_H} \geq \frac{p_{LL}}{p_{HH}-p_{HL}}$,

b) $\frac{\Delta_L}{\Delta_H} \geq \frac{p_{LH}-p_{LL}}{p_{HH}-p_{HL}}$,

c) $\Delta_H \geq 0$.

Proof of Proposition 13. The function $\mathcal{C}(P)$ is, in the case of a capped bonus:

$$\begin{aligned} c_{HH}^* &= W + b + r_0 p_{HH} \\ c_{HL}^* &= W + b + r_0 p_{HL} \\ c_{LH}^* &= W + b + r_0 p_{HL} \\ c_{LL}^* &= W + r_0 p_{LL}, \end{aligned}$$

and hence,

$$\begin{bmatrix} \Delta_H \\ \Delta_L \\ c_{LL}^* \end{bmatrix} = \begin{bmatrix} 0 & 0 & \alpha_H \\ 0 & 1 & \alpha_L \\ 1 & 0 & p_{LL} \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \end{bmatrix}.$$

The solution to this system is

$$\begin{aligned} W &= c_{LL}^* - p_{LL}\frac{\Delta_H}{\alpha_H}, \\ b &= \Delta_L - \alpha_L \frac{\Delta_H}{\alpha_H}, \\ r_0 &= \frac{\Delta_H}{\alpha_H}. \end{aligned}$$

■

Proof of Proposition 10 (in text). (Necessity) The case $\pi_A = \pi_B$ implies no variation in stock prices, so ad–hoc payments are needed to implement any variation in consumption. Provided $\pi_A \neq \pi_B$, for a firm of general type, we have, for a capped bonus and $p_{LH} > p_0$, the system becomes:

$$\begin{aligned} c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\ \Delta_H &= r_0 \alpha_H + s_0 \alpha_H + r_H \alpha_H + s_H (p_{HH} - p_H) \\ \Delta_L &= b + r_0 \alpha_L + s_0 (p_{HL} - p_0) + r_L \alpha_L + s_L (p_{HL} - p_L) \\ 0 &= r_H p_{HL} - r_L p_{HL} - s_L (p_{HL} - p_L). \end{aligned}$$

and it is clear that monotonicity follows from any combination of instruments. This is trivially also true for the case $p_{LH} \leq p_0$, and for a linear bonus. (Sufficiency) The following solution implements any optimal consumption for a general–type firm (the same we used in the proof of Proposition 4 for a type $\mathcal{H}$ firm):

$$\begin{aligned} W &= c_{LL} \\ b &= \Delta_L \\ s_H &= \frac{\Delta_H}{p_{HH} - p_H} \\ s_0 &= r_0 = r_L = r_H = s_L = 0. \end{aligned}$$

■

References

[1] Acharya, Viral V., Kose John, and Rangarajan K. Sundaram. 2000. "Contract Renegotiation and the Optimality of Resetting Executive Stock Options." *Journal of Financial Economics* 57: 65–101.

[2] Jorge G. Aseff and Manuel S. Santos, "Stock options and managerial optimal contracts," *Economic Theory* 26, 813–837 (2005)

[3] Bebchuk, Lucian A., Alma Cohen, and Holger Spamann, 2010, "The wages of failure: Executive compensation at Bear Stearns and Lehman 2000-2008," *Yale Journal on Regulation* 27, 257–282.

[4] Bolton, P., Hamid Mehran, and Joel Shapiro. "Executive Compensation and Risk Taking", Mimeo (2010).

[5] Brenner, M., R. Sundaram and D. Yermak, "Altering the terms of executive stock options," *Journal of Financial Economics* **57**, (2000) 103-128

[6] Carter, M. E., L. J. Lynch and I. Tuna (2007), 'The Role of Accounting in the Design of CEO Equity Compensation', *The Accounting Review*, Vol. 82, No. 2, pp. 327–57.

[7] Carter, Mary Ellen, and Luann J. Lynch. 2001. "An Examination of Executive Stock Option Repricing." *Journal of Financial Economics* 61 (August): 207–25.

[8] Celentani, M. and R. Loveira, "A Simple Explanation of the Relative Performance Evaluation Puzzle," *Review of Economic Dynamics*, vol. 9(3), pages 525-540, July (2006).

[9] Chance, D., R. Kumar, R. Todd, "The 'repricing' of executive stock options". *Journal of Financial Economics* **57** 129-154 (2000)

[10] Chen, M.A. "Executive Option Repricing, Incentives, and Retention". *The Journal of Finance*, Vol. 59, No. 3, pp. 1167-1199 (2004)

[11] Clementi, Gian Luca, and Thomas F. Cooley, 2009. "Executive Compensation: Facts," NBER Working Papers 15426, National Bureau of Economic Research, Inc.

[12] Clementi, Gian Luca, Cooley, Thomas F. and Wang, Cheng, 2006. "Stock grants as a commitment device," *Journal of Economic Dynamics and Control*, Elsevier, vol. 30(11), pages 2191-2216, November.

[13] Chance, D. M.; R. Kumar; and R. B. Todd. "The 'Repricing' of Executive Stock Options." *Journal of Financial Economics* 57 (2000): 129–54.

[14] Core, J. E., and W. R. Guay. "The Use of Equity Grants to Manage Optimal Equity Incentive Levels." *Journal of Accounting and Economics* 28 (1999): 151–84.

[15] Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, Yuliy Sannikov, "Dynamic CEO Compensation," *The Journal of Finance*, 2012, vol. 67(5), p. 1603-1647

[16] Edmans, Alex, and Q. Liu, "Inside Debt", *Review of Finance* (2010), 1–28

[17] Fahlenbrach, Rudiger, and Rene M. Stulz, 2009, "Managerial ownership dynamics and firm value," *Journal of Financial Economics* 92, 342–361.

[18] Gabaix, X. and Augustin Landier, "Why Has CEO Pay Increased So Much?", *Quarterly Journal of Economics*, vol. 123(1), 2008, p. 49-100.

[19] Gillan, Stuart L., Jay C. Hartzell and Robert Parrino. "Explicit versus Implicit Contracts: Evidence from CEO Employment Agreements." *The Journal of Finance*, Volume 64, Issue 4, pages 1629–1655, August 2009

[20] Grossman, Sanford and Oliver D. Hart. "An Analysis of the Principal–Agent Problem." *Econometrica* **51**, Issue 1 (Jan.,1983), 7-46.

[21] Hall, Brian. "The Design of Multi-Year Option Plans." Journal of Applied Corporate Finance 12, no. 2 (summer 1999): 97–106.

[22] Hall, Brian J., and Jeffrey B. Liebman. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* (August 1998): 653–691.

[23] Hall, Brian J. and Thomas A. Knox. "Underwater Options and the Dynamics of Executive Pay-to-Performance Sensitivities," *Journal of Accounting Research*, Blackwell Publishing, vol. 42(2), pages 365-412, 05. (2004)

[24] Holmström, B. "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. **10** (1) pp. 74-91. (1979)

[25] Jarque, Arantxa, and Brian Gaines. "Regulation and the Composition of CEO Pay," FRB Richmond Economic Quarterly, Vol. 98(4), Fourth Quarter 2012, p. 309-348.

[26] Kadan, O. and J. Swinkels. “Stocks or Options? Moral Hazard, Firm Viability, and the Design of Compensation.” *Review of Financial Studies* (2007)

[27] Kole, Stacey R., 1997, “The complexity of compensation contracts,” *Journal of Financial Economics*, 43, 79–104.

[28] Miller, Nolan.“Moral Hazard with Persistence and Learning”, Manuscript. (1999)

[29] Murphy, Kevin J. 1999. “Executive Compensation.” In Handbook of Labor Economics, edited by Orley Ashenfelter and David Card. NewYork: Elsevier Science North Holland, 2485–563.

[30] Wang, Cheng (1997), “Incentives, CEO Compensation and Shareholder Wealth in a Dynamic Agency Model,” *Journal of Economic Theory* 76: 72-105.